

Human immune syste
Nucleotide sequenc
Tumour suppressor
Human immune syste
Human immune syste
Plasmodium falcipa
Human AKRP10 gene
Human AKRP10 gene
Staphylococcus aur
Human immune syste
Human gene regulat
Human immune syste
BAC contigling rep
Arabidopsis thalia
Staphylococcus aur
Human immune syste
Human immune syste
Human immune syste
Human immune/hema
Human immune/hema
Human immune/hema
Human immune/hema
Human immune syste
Human immune syste
Human immune/hema
Human immune/hema
Human genomic DNA
Human DNA for a no
Human muscle/skele
Tumour suppressor
Human cDNA clone (
Drosophila melanog
Borrelia burgdorfe
Staphylococcus aur
Human immune syste
Drosophila melanog
Chromosome 13q21-q
Human metastasis 13q21-q

sequence 315 BP; 108 A; 58 C; 49 G; 100 T; 0 other;

```

conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

| RESULT | 2 |
|----------|-------------------------------------|
| AAV21209 | |
| ID | AAV21209 standard; DNA; 1664976 BP. |
| XX | |

Complete genome sequence of methano-genic archaeon, *Methanococcus jannaschii* - useful in identification of *M. jannaschii* genome

The present sequence represents the complete 1.66-megabase pair genome sequence of the *Methanococcus jannaschii* circular chromosome. The present invention describes *M. jannaschii* open reading frames from the genome sequence. The invention also describes a computer based system for identifying fragments of the *M. jannaschii* genome that are homologous to target nucleotide sequences, comprising: (a) data storage bp sequence (see AAYV1209, AAYV1210 and AAYV1211), or a nucleotide sequence at least 99.9% identical to it; (b) search means for comparing a target sequence to the nucleotide sequence of the data storage means to identify a homologous sequence, and (c) retrieval means for obtaining the homologous sequence. The method, which is based on whole genome random sequencing of an autotrophic archaeon *M. jannaschii*, the genome of which consists of 3 physically distinct elements, a large circular chromosome (the 1664976 bp sequence given in AAYV1209), a large circular extra-chromosomal element (the 58407 bp sequence given in AAYV1210), and a small circular extra-chromosomal element (the 16550 bp sequence given in AAYV1211), can be used in the identification of *M. jannaschii* genome fragments.

| | | | | |
|--------------------------|--------|----------------|-----------|-----------------|
| Query Match | 12.1%; | Score 38; | DB 19; | Length 1664976; |
| Best Local Similarity | 54.2%; | Pred. No. 5.1; | | |
| Matches 77; Conservative | 0; | Mismatches 65; | Indels 0; | Care 0 |

| | |
|------------|----------------------------------|
| RESULT | 3 |
| ABL12244/c | |
| ID | ABL12244 standard; cDNA; 5516 BP |
| XX | |

| | |
|----|-----------------------------|
| PF | 23-MAR-2001; 2001WO-US09231 |
| XX | |
| PR | 23-MAR-2000; 2000US-191637P |
| PR | 11-JUL-2000; 2000US-0614150 |
| XY | |

AA (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, LI FMD, Myers EW,
 PI
 XX WPI; 2001-656860/75.
 DR

PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 XX
 PS Claim 1; SEQ ID No 1; 192pp; French.

XX The present sequence is the genome sequence of Listeria monocytogenes
 CC ECP-e. This sequence and fragments of this sequence are useful for
 CC selecting probes and primers for detecting genes in L. monocytogenes
 CC and related organisms, and to study genetic polymorphisms and
 CC proteins (AB84/7297-AB85/0149) expressed from the present sequence are
 CC useful for raising specific antibodies, identification of L.
 CC monocytogenes and related organisms, and for biosynthesis and
 CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
 CC proteins encoded by it are also useful for selecting compounds that
 CC regulate gene expression and cell replication and modulate L.
 CC monocytogenes-related diseases. In addition, this sequence and proteins
 CC encoded by it are useful in pharmaceutical and vaccines compositions for
 CC the treatment or prevention of infections by L. monocytogenes and related
 CC organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Query Match 11.2%; Score 35.2; DB 24; Length 2944528;
 Best Local Similarity 60.4%; Pred. No. 30;
 Matches 58; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

OY 3 aaccgtgactgaattttagaagaataactttgagaatagaacaataatgaatcgttc 62
 DB 338651 aacaaatagaataatttcaacaaataactagcccaaatataacaaagaagaatgtcgtalt 338710
 OY 63 tccacaccttaagtatactccttagagatcaccg 98
 DB 338711 cgaaccctccagaagaagctgagtgatagaagctcgc 338746

RESULT 6
 AA158172
 ID AA158172 standard; cDNA; 2597 BP.
 XX
 AC AA158172;

XX 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 375.

XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW leukæmia; ss.

XX Homo sapiens.

PN WO200153312-A1.

XX 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEO INC.
 XX
 PI Tang YI, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou F, Goodrich R, Drmanac R;
 XX
 DR WPI: 2001-442253/47.
 P-PSDB; AAM39016.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 375; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161365) and
 CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC assays for receptor activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SO Sequence 2597 BP; 733 A; 607 C; 690 G; 567 T; 0 other;

Query Match 11.1%; Score 35; DB 22; Length 2597;
 Best Local Similarity 55.3%; Pred. No. 8;
 Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

OY 148 caccctatcatcaagtagtgatgactacaatttgaataatagatgacacataaac 207

DB 1434 catcttcgttgagtttagaagacgagatcgatgacacatgacctgtcctcaattgaa 1493

OY 208 tggagtttagaacaacagtagtagaaggaataacacattccatcccttcaagagatca 267

DB 1494 tcaggtttgccaataacacagtagaagacgaagaattcatcattagcgcacatcc 1553

OY 268 ttt 270

DB 1554 ttt 1556

RESULT 7
 AAAX20253
 ID AAAX20253 standard; DNA; 26811 BP.

XX AAAX20253;

DT 04-MAY-1999 (first entry)

DE Borrelia burgdorferi polynucleotide sequence #6.

XX Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.

OS Borrelia burgdorferi.

PN WO9858943-A1.

PR 30-DEC-1998.

PF 18-JUN-1998; 98WO-US12764.

```

PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI White OR;
XX
DR WPI; 1999-081217/07.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
XX
PS Claim 1; Page 851-867; 1128pp; English.
XX
CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
XX
SQ Sequence 26811 BP; 8978 A; 3456 C; 3596 G; 10780 T; 1 other;

Query Match 11.1%; Score 35; DB 20; Length 26811;
Best Local Similarity 49.7%; Pred. No. 13;
Matches 89; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 122 tcagtggtgtgccttggtagccacccattatccaagtagagtagtactcaattt 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 14519 taacttttatacaattttatacaagaacaccccttaacagttttataaattt 14578

QY 182 tgaataatgatgtgcacacaataaactggagttatggaacatcagtagaagaataac 241
   || |||| ||||| ||||| ||||| ||||| ||||| |||||
DB 14579 tgccataattatataataataattccattcaacttaacgttataaagcacttaa 14638

QY 242 aacattccaccccttaccagagatcatttactgcacacccagagataattgtcagtgt 300
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 14639 attactctaaataaataatgacatttaccatttaataaagcaataattgactttatat 14697

RESULT 8
ABU17122
ID ABU17122 standard; DNA; 4084 BP.
XX
AC ABU17122;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 2839.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX

```

```

PA (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 2839; 21pp + Sequence Listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins
CC (ABU57737-ABU72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 4084 BP; 1059 A; 913 C; 872 G; 1240 T; 0 other;

Query Match 11.0%; Score 34.8; DB 23; Length 4084;
Best Local Similarity 48.1%; Pred. No. 10;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 83 cttagagatctacagcctcccttagggagacatacaagtcggtgtgtgtccttgtg 142
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3493 catcaatacatatacatatacatatgtaacataatgtaaacattatctcgttgaag 3552

QY 143 agtccacccattatcaagttagtgcacaaatttggaaatagattgcacaca 202
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3553 ctgcacacatttgcagcccaaatgctcatcaattgttacaacattatccaataa 3612

QY 203 taactggaattatggaacatcagtagaaggaataacaacattccatcccttaca 262
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3613 catgtttattatatttctcatttgcacatcaataagaacattgtttgatttca 3672

QY 263 gatcatttactgcacactcagatata 288
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3673 gctttcttctgttactctgtcatga 3698

RESULT 9
ABU32774/C
ID ABU32774 standard; DNA; 15122 BP.
XX
AC ABU32774;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 747.
XX
KW Human; immune system disease; cytosine methylation; antistimatic;
KW antiarteriosclerotic; antianaemic; cytosolic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
XX
OS
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX

```


DR WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 XX
 PS Claim 1; Page 291-611; 1912pp; English.
 XX
 CC The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AYA34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC AYA34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;
 Query Match 11.0%; Score 34.8; DB 20; Length 1230025;
 Best Local Similarity 52.0%; Pred. No. 33;
 Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 QY 70 ctaagatatacctctagagatctacagcctcccttaaggagacatacaagtcagtgt 129
 Db 1072911 CTTAAGTTAGTAAATCTCTATCTAATCACTGCTTCTGTAATCCCTTAAGAAAGAGCT 1072852
 QY 130 gttgccttcttgagtcaccactataatcaagtagtagtactacaatttgaataa 189
 Db 1072851 GTTGTATTTTAAAGTCTTTTAAACAACACTTGAGTTCAGGCCAAACAGTAGAAGA 1072792
 QY 190 gattgcacacataaactgagttatg 219
 Db 1072791 AATTTCTCTCAAAAAAATTTGTTATAG 1072762
 RESULT 12
 AAS46370/C
 ID AAS46370 standard; DNA; 6169 BP.
 AC AAS46370;
 XX
 DT 18-DEC-2001 (first entry)
 DE Tumour suppressor gene derived chemically modified sequence #92.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 PA (EPig-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602752/68.
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for

PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 PS Claim 1; SEQ ID No 92; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (cp DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6169 BP; 1840 A; 112 C; 1192 G; 3025 T; 0 other;
 Query Match 10.9%; Score 34.2; DB 22; Length 6169;
 Best Local Similarity 54.3%; Pred. No. 16;
 Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
 QY 187 atgattgtcacacaataaactgaggttatggaacaatcagaaggaatacaacat 246
 Db 2251 AAACATTCTAATTAATTAATCCAAATCAAAAAAATTTATATCAATAAAAAAGCTT 2292
 QY 247 tccatcccttacagagatcattactgcaactcagataattgtcatggtatc 306
 Db 2291 ACATTTCCCTTACAAATATATATTAATAAATACTAAAAAATTTTATCTATATTTTAAC 2232
 QY 307 tacttat 313
 Db 2231 TTTTAAT 2225
 RESULT 13
 ABL3323/C
 ID ABL3323 standard; DNA; 9741 BP.
 XX
 AC ABL3323;
 XX
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 1296.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; antiarthritic; antidiabetic; antipsoptic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX

PR 30-JUN-2000; 2000DE-103259.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPiG-) EPIGENOMICS AG.
XX
XX Olek A., Piepenbrock C., Berlin K;
XX
DR WPI.; 2002-130909/17.
XX
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS
XX Claim 1; SEQ ID NO 1676; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

SO
Sequence 18624 BP; 5682 A; 113 C; 3466 G; 9361 T; 2 other;

Query Match 10.9%; Score 34.2; DB 24; Length 18624;
Best Local Similarity 47.9%; Pred. No. 20;
Matches 128; Conservative 128; Mismatches 128; Indels 1

[illegible]

| | | |
|----|----------|---|
| XX | PN | EP447956-A. |
| XX | FT | /note="N-terminal" |
| XX | FT | /label="Sporozoite-antigen |
| XX | FT | /*tag= a |
| XX | FT | 948..3086 |
| XX | FT | Location/Qualifiers |
| XX | OS | Plasmodium falciparum. |
| XX | XX | Vaccine; NXY gene; malaria; antibodies; ss. |
| XX | XX | Plasmodium falciparum sporozoite antigen N-terminal |
| XX | DE | |
| XX | DT | 05-DEC-1991 (first entry) |
| XX | AC | AAQ13727; |
| XX | AAQ13727 | standard; DNA; 3086 BP. |
| XX | ID | AAQ13727 |
| XX | RESULT | 15 |

XX 25-SEP-1991.
 PD 14-MAR-1991; 91EP-0103920.
 XX 23-MAR-1990; 90CH-0000970.
 PF (HOFF) HOFFMANN-LA ROCHE AG.
 XX Certa U, Guttinger M, Matile H;
 XX WPT: 1991-28289/39.
 DR P-PSDB: AARI3991.
 XX
 XX New antimalarial polypeptide(s) - corresp. to specific epitope(s)
 PT of the Plasmodium falciparum sporozoite antigen, and are useful
 PT as vaccines or to prepare antigens
 XX
 PS Disclosure: Fig 2; 36pp; German.
 XX
 CC The sequence is useful in the prodn. of polypeptides corresp. in
 CC at least one specific epitope with the P. falciparum sporozoite
 CC antigen N-terminal. These polypeptides are useful in the prodn.
 CC of antimalarial vaccines.
 CC The gene, designated, NXY, was isolated from a genebank produced
 CC from P. falciparum cells (K1 isolate) screened with antiserum
 CC obtained by immunising rabbits with P. falciparum isolate NF54.
 CC See also AA013728.
 XX
 SQ Sequence 3088 BP; 1375 A; 273 C; 575 G; 865 T; 0 other;

Query Match 10.8%; Score 34; DB 12; Length 3088;
 Best Local Similarity 61.1%; Pred. No. 15;
 Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 OY 154 atattcaagtaggtactacaatttgaatatagattgtcacacaataaactgagat 213
 DB 426 ataataaacaataagaanaaactaataatgataataatacattgaacaataatcgcgaant 485
 OY 214 ttatggaacatcagtagagaagaatacaa 243
 DB 486 ctaggggcacatctgtacagaagcgcacaaa 515

Search completed: September 16, 2002, 03:05:06
 Job time: 7984 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:05:06 : Search time 390.12 Seconds
(without alignments)
3938.887 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895
Sequence: 1 ctaatctgtctagctaacgc.....aataatatttttaaatgac 895

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802:*

1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|-------|--------------------|------|-------------|
| 1 | 895 | 100.0 | 895 | 21 AA29726 |
| 2 | 895 | 100.0 | 895 | 22 AAH77949 |
| 3 | 895 | 100.0 | 1171 | 22 AAH77951 |
| 4 | 858 | 95.9 | 870 | 21 AA298202 |
| 5 | 856.4 | 95.7 | 1148 | 22 ABA09257 |
| 6 | 852 | 95.2 | 870 | 21 AA265001 |
| 7 | 852 | 95.2 | 870 | 22 AAS45994 |
| 8 | 852 | 95.2 | 870 | 22 AAF92070 |
| 9 | 852 | 95.2 | 870 | 22 AAF44147 |

| | | | | | | |
|----|-------|------|-------|----|----------|--------------------|
| 10 | 699 | 78.1 | 849 | 21 | AAA08343 | Human TGC-440 secr |
| 11 | 630 | 70.4 | 654 | 21 | AAC59829 | Human secreted pro |
| 12 | 525.8 | 58.7 | 533 | 22 | AA668195 | Human lung tumour |
| 13 | 445.2 | 49.7 | 511 | 19 | AAV68618 | Human secreted pro |
| 14 | 362.4 | 40.5 | 365 | 20 | AAK40454 | Human secreted pro |
| 15 | 357 | 39.9 | 357 | 21 | AAA08344 | Human TGC-440 secr |
| 16 | 312 | 34.9 | 312 | 22 | AA668125 | Human lung tumour |
| 17 | 291 | 32.5 | 291 | 21 | AAA08345 | Mature human TGC-4 |
| 18 | 262.2 | 29.3 | 764 | 21 | AAA08349 | Mouse TGC-440 secr |
| 19 | 245.6 | 27.4 | 357 | 21 | AAA08350 | Mouse TGC-440 secr |
| 20 | 241.8 | 27.0 | 959 | 21 | AAA08346 | Rat TGC-440 secret |
| 21 | 228.4 | 25.5 | 235 | 20 | AAK40453 | Human secreted pro |
| 22 | 228 | 25.5 | 357 | 21 | AAA08347 | Rat TGC-440 secret |
| 23 | 200.4 | 22.4 | 291 | 21 | AAA08351 | Mature mouse TGC-4 |
| 24 | 189.2 | 21.1 | 291 | 21 | AAA08348 | Mature rat TGC-440 |
| 25 | 189 | 21.1 | 195 | 16 | AAT25820 | Human gene signatu |
| 26 | 73.4 | 8.2 | 97 | 22 | AA507713 | Cervical cancer pr |
| 27 | 45.8 | 5.1 | 518 | 22 | AAH70020 | Human cervical can |
| 28 | 44.6 | 5.0 | 51 | 21 | AAA08356 | Human TGC-440 secr |
| 29 | 42.8 | 4.8 | 456 | 22 | AAH71445 | Human cervical can |
| 30 | 39 | 4.4 | 775 | 22 | AAK71873 | Human immune/haema |
| 31 | 39 | 4.4 | 775 | 22 | AAK71874 | Human immune/haema |
| 32 | 38.6 | 4.3 | 356 | 22 | AAK58069 | Human immune/haema |
| 33 | 36.4 | 4.1 | 704 | 22 | AAL13204 | Human breast cance |
| 34 | 36 | 4.0 | 48452 | 23 | ABL07108 | Drosophila melanog |
| 35 | 35.8 | 4.0 | 47 | 21 | AAA08355 | Human TGC-440 secr |
| 36 | 35.6 | 4.0 | 632 | 22 | AAH31453 | Human olfactory re |
| 37 | 35.6 | 4.0 | 6792 | 22 | AAK52133 | Human polynucleoti |
| 38 | 35.6 | 4.0 | 6815 | 22 | AAH90035 | Human bone marrow |
| 39 | 35.6 | 4.0 | 7317 | 22 | AAH89922 | Human bone marrow |
| 40 | 35.6 | 4.0 | 51469 | 22 | AAK69322 | Human immune/haema |
| 41 | 35.6 | 4.0 | 51469 | 22 | AAK70270 | Human immune/haema |
| 42 | 35.6 | 4.0 | 51469 | 22 | AAK78813 | Human immune/haema |
| 43 | 35.4 | 4.0 | 2484 | 23 | AAK85729 | DNA encoding novel |
| 44 | 35.4 | 4.0 | 7642 | 24 | ABL33116 | Human immune syste |
| 45 | 35 | 3.9 | 1610 | 22 | AAI93220 | Human polynucleoti |

ALIGNMENTS

RESULT 1

AA29726

ID AA29726 standard; DNA: 895 BP.

AC AA29726:

XX 27-MAR-2000 (first entry)

AC 27-MAR-2000 (first entry)

XX 27-MAR-2000 (first entry)

DE Human Lung specific gene-3.

XX Human Lung specific gene-3.

KW Lung Specific Gene; LSG; human; diagnostic marker;

KW prognosticate; Lung cancer; diagnosis; ds.

OS Homo sapiens.

XX Homo sapiens.

XX W09960160-A1.

XX W09960160-A1.

XX 25-NOV-1999.

XX 25-NOV-1999.

XX 12-MAY-1999; 99MO-US10344.

XX 12-MAY-1999; 99MO-US10344.

XX 21-MAY-1998; 98US-0086212.

XX 21-MAY-1998; 98US-0086212.

XX (DIAD-) DIADEXUS LLC.

XX (DIAD-) DIADEXUS LLC.

XX Yang F, Macina RA, Sun Y;

XX Yang F, Macina RA, Sun Y;

XX WPI; 2000-116320/10.

XX WPI; 2000-116320/10.

XX A new method for diagnosing, monitoring and staging lung cancer

XX A new method for diagnosing, monitoring and staging lung cancer

XX Example 1; Pages 35-36; 40pp; English.

PS Example 1; Pages 35-36; 40pp; English.

XX The present sequence is a lung specific gene (LSG) from human
 CC clone ID 126263. The LSG has high level of tissue specificity for lungs
 CC and is overexpressed in cancerous tissues. The sequence serves as a
 CC diagnostic marker for detecting, monitoring, staging and prognosticating
 CC lung cancer. The diagnosis involves comparing levels of LSG in samples
 CC obtained from patient and normal control.
 XX
 SQ Sequence 895 BP; 245 A; 250 C; 183 G; 217 T; 0 other;

Query Match 100.0%; Score 895; DB 21; Length 895;
 Best Local Similarity 100.0%; Pred. No. 2e-270;
 Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctaactgttactgaagcaagagcgtacccctcctgttctcgcctccaatgga 60
 Db 1 ctaactgttactgaagcaagagcgtacccctcctgttctcgcctccaatgga 60
 OY 61 acgtctgctggactaaagacacacacagcgtatcctcgtacatgcatccc 120
 Db 61 acgtctgctggactaaagacacacacagcgtatcctcgtacatgcatccc 120
 OY 121 caggatcaagagcctccagcagggacccctccattatctctccaacttacagct 180
 Db 121 caggatcaagagcctccagcagggacccctccattatctctccaacttacagct 180
 OY 181 gcaacgaagttgcatgaataagttcattctcctccctccctgctgctccacta 240
 Db 181 gcaacgaagttgcatgaataagttcattctcctccctccctgctgctccacta 240
 OY 241 ctgattccatggtctctcagcagcctgaatccagggtcgcagagagccagagc 300
 Db 241 ctgattccatggtctctcagcagcctgaatccagggtcgcagagagccagagc 300
 OY 301 ggcagagcttctagaagatgctccagaaagcgccaaagatgtgagtcgaagatt 360
 Db 301 ggcagagcttctagaagatgctccagaaagcgccaaagatgtgagtcgaagatt 360
 OY 361 ttctctgagagcccgagaaataatcagcaggtctcgtggtgcgcaagaagcagtc 420
 Db 361 ttctctgagagcccgagaaataatcagcaggtctcgtggtgcgcaagaagcagtc 420
 OY 421 cccgtgatcaatcctcagaggaatgtaagaacaagacacacaaaggcaccagaag 480
 Db 421 cccgtgatcaatcctcagaggaatgtaagaacaagacacacaaaggcaccagaag 480
 OY 481 ccaacaagacatccagagcctgcagcaatcttccaacaatgtcagctaaagctt 540
 Db 481 ccaacaagacatccagagcctgcagcaatcttccaacaatgtcagctaaagctt 540
 OY 541 gctctgcttctgtagagcctcagcagccactctccaatlaaacatctcagcca 600
 Db 541 gctctgcttctgtagagcctcagcagccactctccaatlaaacatctcagcca 600
 OY 601 agacagtggagcacacccacagacactctctccaccactacctccactgtacc 660
 Db 601 agacagtggagcacacccacagacactctctccaccactacctccactgtacc 660
 OY 661 acccctaataatccagtgctctcaaaaagcagtgtttccaagatcaattttgttg 720
 Db 661 acccctaataatccagtgctctcaaaaagcagtgtttccaagatcaattttgttg 720
 OY 721 cttctctcctcctctctctcctcgtcagctcctagcctggtccctccctaccag 780
 Db 721 cttctctcctcctctctctcctcgtcagctcctagcctggtccctccctaccag 780
 OY 781 taagcttaataatcctaaagattccaggaactgagcttccctcagctagtgatcatt 840
 Db 781 taagcttaataatcctaaagattccaggaactgagcttccctcagctagtgatcatt 840
 OY 841 cttaaatgcaatcaggaagtagcaacaggaagtcataataatttttaaatgtc 895
 Db 841 cttaaatgcaatcaggaagtagcaacaggaagtcataataatttttaaatgtc 895

Db 841 cttaaatgcaatcaggaagtagcaacaggaagtcataataatttttaaatgtc 895

RESULT 2
 ID AAH77949
 AAH77949 standard; DNA; 895 BP.

AC AAH77949;
 DT 13-NOV-2001 (first entry)
 DE Nucleotide sequence of a human Lng104 polypeptide.
 XX Human; lung cancer specific gene; LSG; Lng104; lung cancer; ss.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT CDS 196..555
 FT /tag= a
 FT /product= "Lng104"
 XX
 PN WO200161055-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001MO-US05674.
 XX
 PR 17-FEB-2000; 2000US-0183188.
 XX
 PA (DIAD-) DIADEXUS INC.
 XX
 PI Chen S, Sun Y, Macina RA;
 DR WPI: 2001-529917/58.
 DR P-PSDB; AAG63977.
 PT New lung cancer specific gene for the treatment and diagnosis of lung
 PT cancer -
 PS
 PS Claim 1; Page 113; 119pp; English.
 XX
 XX
 CC The present sequence represents a human lung cancer specific gene
 CC (LSG), and encodes a polypeptide designated Lng104. LSGs are useful
 CC in the treatment and diagnosis of lung cancer. The treatment of lung
 CC cancer comprises the administration of a molecule which down regulates
 CC the expression of an LSG. An immune response can be mounted against a
 CC target cell expressing an LSG. Identification of potential therapeutic
 CC agents for use in imaging and treating lung cancer which comprises
 CC screening molecules for an ability to bind to or decrease expression
 CC of an LSG relative to LSG in the absence of the agent where the ability
 CC of a molecule to bind to the LSG or decrease expression of the LSG is
 CC indicative of the molecule being useful in imaging and treating lung
 CC cancer.
 XX
 SQ Sequence 895 BP; 245 A; 250 C; 183 G; 217 T; 0 other;

Query Match 100.0%; Score 895; DB 22; Length 895;
 Best Local Similarity 100.0%; Pred. No. 2e-270;
 Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctaactgttactgaagcaagagcgtacccctcctgttctcgcctccaatgga 60
 Db 1 ctaactgttactgaagcaagagcgtacccctcctgttctcgcctccaatgga 60
 OY 61 acgtctgctggactaaagacacacacagcgtatcctcgtacatgcatccc 120
 Db 61 acgtctgctggactaaagacacacacagcgtatcctcgtacatgcatccc 120
 OY 121 caggatcaagagcctccagcagggacccctccattatctctccaacttacagct 180
 Db 121 caggatcaagagcctccagcagggacccctccattatctctccaacttacagct 180

```
OY 181 gcaaccagatgtcgatgaaagtctcaatctctctcctcctcctgtgtgtgcactaaag 240
    |||||||
DB 181 gcaaccagatgtcgatgaaagtctcaatctctcctcctcctcctgtgtgtgcactaaag 240
OY 241 ctgatgtccatgtgtctctagcagctgtaaatccagggtgtgcagagcgacagagcgca 300
    |||||||
DB 241 ctgatgtccatgtgtctctagcagctgtaaatccagggtgtgcagagcgacagagcgca 300
OY 301 ggcacgagcttctagagatggtctccaggagcgcccaagaatgtgagtgcacaagtgtg 360
    |||||||
DB 301 ggcacgagcttctagagatggtctccaggagcgcccaagaatgtgagtgcacaagtgtg 360
OY 361 ttccctgagagcccgagaaagaatattcatgacatgtctgggctgtgcacaagaagcagttc 420
    |||||||
DB 361 ttccctgagagcccgagaaagaatattcatgacatgtctgggctgtgcacaagaagcagttc 420
OY 421 cccctgtgacatcttcaaggcgcaatgtgaaagaaacacacaaaggcgacacagaaag 480
    |||||||
DB 421 cccctgtgacatcttcaaggcgcaatgtgaaagaaacacacaaaggcgacacagaaag 480
OY 481 ccaaaacagcatctcagagcctgcagcaattcttcaacaatgtcagctaaagaactt 540
    |||||||
DB 481 ccaaaacagcatctcagagcctgcagcaattcttcaacaatgtcagctaaagaactt 540
OY 541 gctctgcttctgtagagctctgagcgccactcttccaatlaaacattctcagccaaga 600
    |||||||
DB 541 gctctgcttctgtagagctctgagcgccactcttccaatlaaacattctcagccaaga 600
OY 601 agacagtgagcaacactacagacactcttcttccacactcctcccaactgtaccc 660
    |||||||
DB 601 agacagtgagcaacactacagacactcttcttccacactcctcccaactgtaccc 660
OY 661 acccccaatctctcagcgtctcaaaagatgttttcaagatcatttgttgttg 720
    |||||||
DB 661 acccccaatctctcagcgtctcaaaagatgttttcaagatcatttgttgttg 720
OY 721 cctctcctagtgctctctcctcgtcagctcttaagcctgtgcctcccttaccagagct 780
    |||||||
DB 721 cctctcctagtgctctctcctcgtcagctcttaagcctgtgcctcccttaccagagct 780
OY 781 taggcttaattcaactgaagaattccaggaaactgtagcttccctagtagtgcatttaac 840
    |||||||
DB 781 taggcttaattcaactgaagaattccaggaaactgtagcttccctagtagtgcatttaac 840
OY 841 cttaaatgcaatcaggaagaatgacaaacagaagtcataatatttttaaatgtc 895
    |||||||
DB 841 cttaaatgcaatcaggaagaatgacaaacagaagtcataatatttttaaatgtc 895

RESULT 3
AAH77951
ID AAH77951 standard; DNA; 1171 BP.
XX
XX
AC AAH77951;
XX
XX
DT 13-NOV-2001 (first entry)
XX
DE Nucleotide sequence of a human Lng104 polypeptide.
XX
KW Human: lung cancer specific gene; LSG; Lng104; lung cancer; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS 217..576 Location/Qualifiers
FT /tag= a
FT /product= "Lng104"
XX
XX
PN WO200161055-A2.
XX
PD 23-AUG-2001.
XX
```

```
PF 20-FEB-2001; 2001WO-US05674.
XX
PR 17-FEB-2000; 2000US-0183188.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Chen S, Sun Y, Macina RA;
XX
XX WPI: 2001-529917/58.
XX
XX P-PSDB; AAG63977.
XX
XX New lung cancer specific gene for the treatment and diagnosis of lung
XX cancer -
XX
XX Claim 1; Page 117; 119pp; English.
XX
XX PS
XX CC The present sequence represents a human lung cancer specific gene
XX CC (LSG), and encodes a polypeptide designated Lng104. LSGs are useful
XX CC in the treatment and diagnosis of lung cancer. The treatment of lung
XX CC cancer comprises the administration of a molecule which down regulates
XX CC the expression of an LSG. An immune response can be mounted against a
XX CC target cell expressing an LSG. Identification of potential therapeutic
XX CC agents for use in imaging and treating lung cancer which comprises
XX CC screening molecules for an ability to bind to or decrease expression
XX CC of an LSG relative to LSG in the absence of the agent where the ability
XX CC of a molecule to bind to the LSG or decrease expression of the LSG is
XX CC indicative of the molecule being useful in imaging and treating lung
XX CC cancer.
XX
SQ Sequence 1171 BP; 333 A; 320 C; 225 G; 293 T; 0 other;

Query Match 100.0%; Score 895; DB 22; Length 1171;
Best Local Similarity 100.0%; Pred. No. 2.3e-270;
Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctaatctgtacglaacagcaagacagcgtcaactacactgtctctgcctccaatgga 60
    |||||||
DB 22 ctaatctgtacglaacagcaagacagcgtcaactacactacactgtctctgcctccaatgga 81
OY 61 acgctgagcctggagactaaagcatagacacacagctgagatctcgtacctgaatccc 120
    |||||||
DB 82 acgctgagcctggagactaaagcatagacacacagctgagatctcgtacctgaatccc 141
OY 121 caggagatcaggagcctccagcagggagacatttcattatattctcaagaacttaagct 180
    |||||||
DB 142 caggagatcaggagcctccagcagggagacatttcattatattctcaagaacttaagct 201
OY 181 gcaaccagatgtcgatgaaagtctcaatctctcctcctcctcctgtgtgtgcactaaag 240
    |||||||
DB 202 gcaaccagatgtcgatgaaagtctcaatctctcctcctcctcctgtgtgtgcactaaag 261
OY 241 ctgatgtccatgtgtctctagcagctgtaaatccagggtgtgcagagcgacagagcgca 300
    |||||||
DB 262 ctgatgtccatgtgtctctagcagctgtaaatccagggtgtgcagagcgacagagcgca 321
OY 301 ggcacgagcttctagagatggtctccaggagcgcccaagaatgtgagtgcacaagtgtg 360
    |||||||
DB 322 ggcacgagcttctagagatggtctccaggagcgcccaagaatgtgagtgcacaagtgtg 381
OY 361 ttccctgagagcccgagaaagaatattcatgacatgtctgggctgtgcacaagaagcagttc 420
    |||||||
DB 382 ttccctgagagcccgagaaagaatattcatgacatgtctgggctgtgcacaagaagcagttc 441
OY 421 cccctgtgacatcttcaaggcgcaatgtgaaagaaacacacaaaggcgacacagaaag 480
    |||||||
DB 442 cccctgtgacatcttcaaggcgcaatgtgaaagaaacacacaaaggcgacacagaaag 501
OY 481 ccaaaacagcatctcagagcctgcagcaattcttcaacaatgtcagctaaagaactt 540
    |||||||
DB 502 ccaaaacagcatctcagagcctgcagcaattcttcaacaatgtcagctaaagaactt 561
OY 541 gctctgcttctgtagagctctgagcgccactcttccaatlaaacattctcagccaaga 600
```

human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have antitumour, anti-inflammatory, antimicrobial, neurotrophic, hepatotropic, neuroprotective, cardiovascular and antistatic activities, and can be used in gene therapy. HSPs can be used to treat or prevent disorders associated with decreased activity or function of HSP. Antagonists of HSP are used to treat or prevent disorders associated with increased activity or function of HSP. Such diseases include cell proliferation (including cancer), inflammation, cardiovascular, neurological, reproductive or developmental disorders, (e.g. arteriosclerosis, cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia, asthma, Crohn's disease, microbial or other infections, congestive or ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP nucleic acids can be used for the recombinant production of HSP, for detecting HSP in standard hybridization and amplification assays (for diagnosis and monitoring), in gene therapy, as antisense, triplex-forming or ribozyme therapeutics, for detecting related sequences or genetic variations, and for chromosomal mapping. HSP are also used to raise specific antibodies (Ab) and to screen for agonists and antagonists (potential therapeutic agents). Ab are used to diagnose, or monitor, HSP-related diseases (in usual immunoassays), as therapeutic antagonists, in competitive drug screens, and for purification of HSP from natural sources.

Sequence 870 BP; 244 A; 240 C; 177 G; 209 T; 0 other;

| | | | | |
|-----------------------|--------|---------------------|--------|-------------|
| Query Match | 95.9% | Score 858; | DB 21; | Length 870; |
| Best Local Similarity | 100.0% | Pred. No. 8,2e-259; | | |

```

matches      858;  Conservative    0;  Mismatches    0;  Indels      0;  Gaps      0

```

ggaagcgcggacccaaccaggctg 97

Qy 98 agtactctgaactgagctatctcccaagagatctcagaactccagcaaggaactctcaatta 157
 |||||
 Db 61 agtactctgaactgagctatctcccaagagatctcagaactccagcaaggaactctcaatta 121
 |||||
 Qy 158 tattcttcagaacttcaacactgacacgagcaatttcgataagaaattcaatcttcata 21
 |||||
 Db 121 tattcttcagaacttcaacactgacacgagcaatttcgataagaaattcaatcttcata 180
 |||||

218 tcctcctgtgtgcgcactaatgtgatgttccatggtctcttagcagcctgaatccaagg 277

181 **tcctcctgtgtgctgcactaatgtgatgtlccatggtctcttagcagcctgtaatccaggg** 240

2/8 tcgccagagggccacagggaccgaggtctctagagatgctcaccaggaagcgcc 337

241 tcgccaagagccacaagagacgcaggtctctaagatgtgctccaggaagcgcc 300

338 aaagaatgtctgagctgcaagaattgtgttcctgagagcccgagagaataatcatgacagtgt 397

501 aagdaagcgaagatgtctcctgagagcccgagaagaataatcatgacagtgt 360

ccggccgcgaagcagcgccctgtgatcattccaaggcattgtgaagaaacaa 457

201 ccgagcccaagacagcgcgccctcgtatccattccaaggcgaatgtgaagaacaaca 420

gacacccaagcaccacagaagaagccdaacgaagcatlccagagcctgcagcaatttctca 517

3acacccuagyuaccacagaaagccaacagcaltccagagccctgcagcaattctca 480

|||||aacaaagccagcaagccttcgcccctgtaggagctctgagcgccactctc 577

... aaagccaaagacccccgcgcgcgcggcgcccactctc 540

637

-----cagccacacccacagacacccctctctcc 600

697

Db 601 cactcactctccacgtctaccacccctaatacttccagtgctctcaaaaagcatgtt 660
Qy 698 ttccaagatcatctgttctgtctctctctctctctctctctctctctctctctc 757
Cc |||||||
Db 661 ttccaagatcatctgttctgtctctctctctctctctctctctctctctctc 720
Qy 758 tctgcccccttaccacagcttaggttaattaccctgaagaagattccagaactgtag 817
Db 721 tctgcccccttaccacagcttaggttaattaccctgaagaagattccagaactgtag 780
Qy 818 ctctcctagctagtctcatttaaccttaaatgcacatcagagaagtgcacagaagtc 877
Db 761 ctctcctagctagtctcatttaaccttaaatgcacatcagagaagtgcacagaagtc 840
Qy 878 taatatttttaaatgtc 895
Cc |||||||
Db 841 taaatatttttaaatgtc 858

RESULT 5
ABA09257
ID ABA09257 standard; cDNA: 1148 BP.
XX
AC ABA09257;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human TGC-440 homologue-encoding cDNA, SEQ ID NO:1033.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; anti-inflammatory;
KW antiasthmatic; antiarthritis; haemostatic; antiatherosclerotic;
KW cytosolic; osteopathic; vasculotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antitumor; ss.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
IN 09-AUG-2001.
XX
PD 05-FEB-2001; 2001WO-US03800.
XX
PF 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-457740/49.
DR P-PSDB: ABB12013.
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
PS Claim 1; Page 867-868; 1963pp; English.
XX
CC Sequences ABA10981-AB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which

Cc bind to polypeptides of the invention. Although novel, many of the
Cc polypeptides of the invention have homology to known proteins, thereby
Cc giving an insight into their probable biological activities, and hence
Cc potential therapeutic applications. The polypeptides of the invention may
Cc have various activities, including cytokine, cell proliferation or cell
Cc differentiation activities; stem cell growth factor activity;
Cc haematopoiesis regulatory activity; tissue growth activity;
Cc immunomodulatory activity; activin- or inhibin-related activities;
Cc chemotactic or chemokinetic activities; haemostatic, thrombotic or
Cc thrombolytic activities; receptor or ligand activities; or may be
Cc involved in oncogenesis, cancer cell proliferation or metastasis.
Cc Depending on their biological activities, polypeptides and nucleotides of
Cc the invention are useful for preventing, treating or ameliorating medical
Cc conditions, e.g., by protein or gene therapy. Such conditions include
Cc cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
Cc disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
Cc proliferative retinopathy, atherosclerosis, coronary heart disease,
Cc arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
Cc vascular growth. Polypeptides involved with tissue regeneration and
Cc repair (or nucleic acids encoding them) may be used to promote wound
Cc healing (e.g., of burns, incisions and ulcers), while those with
Cc immunomodulatory activities may be used in the treatment of viral,
Cc bacterial and fungal infections in addition to immune disorders.
Cc Polypeptides with growth factor activity may be used in cell cultures to
Cc promote cell growth. For example, such polypeptides may be used to
Cc manipulate stem cells in culture to give rise to neuroepithelial cells
Cc that can be used to augment or replace cells damaged by illness,
Cc autoimmune disease or accidental damage. The polypeptides and nucleotides
Cc may also be used in the diagnosis of the above conditions, and in drug
Cc screening techniques. The present sequence represents a cDNA encoding a
Cc novel human polypeptide of the invention.
Cc
Cc Sequence 1148 BP: 328 A; 309 C; 222 G; 289 T; 0 other;
Cc
Cc

Query Match 95.7%; Score 856.4; DB 22; Length 1148;
Best Local Similarity 99.9%; Pred. NO. 3e-258;
Matches 857; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 38 cctgtctcgccctcaaatgtagacgtgtgcttggactaaagcatagaccacagctgt 97
Cc |||||||
Db 23 cctgtctcgccctcaaatgtagacgtgtgcttggactaaagcatagaccacagctgt 82
Qy 98 agtaccctgacctgtagtaccctccagagatcagagccctccagaggaacttcacat 157
Cc |||||||
Db 83 agtaccctgacctgtagtaccctccagagatcagagccctccagaggaacttcacat 142
Qy 158 tatcttcaagaacttaacagctgacacagctgtgcagtgaagaattcctactctcc 217
Cc |||||||
Db 143 tatcttcaagaacttaacagctgacacagctgtgcagtgaagaattcctactctcc 202
Qy 218 tctctcctgtgtctgcactaatgtgtatgtccatgtctctctctctctctctctcc 277
Cc |||||||
Db 203 tctctcctgtgtctgcactaatgtgtatgtctctctctctctctctctcc 262
Qy 278 tctgcaagagggccacagggacagggccttctagagatagcttccaggaagcgggc 337
Cc |||||||
Db 263 tctgcaagagggccacagggacagggccttctagagatagcttccaggaagcgggc 322
Qy 338 aagaatgtgagtcacaaagtgttctcttagagcccccagagaagaatatcagagatgt 397
Cc |||||||
Db 323 aagaatgtgagtcacaaagtgttctcttagagcccccagagaagaatatcagagatgt 382
Qy 398 ctgggctgccaaagaagcagtgccctgtgatcatcttcaagggaatgtgaagaacaa 457
Cc |||||||
Db 383 ctgggctgccaaagaagcagtgccctgtgatcatcttcaagggaatgtgaagaacaa 442
Qy 458 gacaccaagagcaccacaggaagccaaacagcatctcagagccctgcagcaattctca 517
Cc |||||||
Db 443 gacaccaagagcaccacaggaagccaaacagcatctcagagccctgcagcaattctca 502
Qy 518 aacaatgcacagtaagaagcttgcctgtgcctttagaagctcgaagccactcttc 577
Cc |||||||

PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 20-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
XX
XX (GENENTECH INC.
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX WPI: 2000-072883/06.
XX P-PSDB: AAY6668.
XX
XX Membrane-bound proteins and related nucleotide sequences -
XX
XX
XX Claim 2; Fig 98; 822pp; English.
XX
XX The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
XX
XX Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 21; Length 870;
Best Local Similarity 100.0%; Pred. No. 6,3e-257;
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgcctcaaatggaacgcctgacctgagactaaagcatagaccacagcctgaatc 103
|||||
DB 1 ctgcgcctcaaatggaacgcctgacctgagactaaagcatagaccacagcctgaatc 60
QY 104 ctgacctgagatcaccacagagatcagagacccacagcaggaaccttcattatct 163
|||||
DB 61 ctgacctgagatcaccacagagatcagagacccacagcaggaaccttcattatct 120
QY 164 tcaagcaactcaagctcagccagcagcagcttgcatgtaaatcttactctccctcc 223
|||||
DB 121 tcaagcaactcaagctcagccagcagcagcttgcatgtaaatcttactctccctcc 180
QY 224 tctgtgtgcacctaagctcagctgctcagctcagcagcagcagcagcagcagc 283
|||||
DB 181 tctgtgtgcacctaagctcagctgctcagctcagcagcagcagcagcagcagc 240
QY 284 gagccacagagagacccagagccagcagctctcagagatgctcagagagcagcagc 343
|||||
DB 241 gagccacagagagacccagagccagcagctctcagagatgctcagagagcagcagc 300
QY 344 gtgagtgcacaaagattggttcctcagagagcccgagagaaatctcatgaagtgtctggc 403
|||||
DB 301 gtgagtgcacaaagattggttcctcagagagcccgagagaaatctcatgaagtgtctggc 360
QY 404 tgcacaaagagcagctgacctgtgactatctcaaggcagcttgagaaacaagacacc 463
|||||
DB 361 tgcacaaagagcagctgacctgtgactatctcaaggcagcttgagaaacaagacacc 420
QY 464 aaagcaccacagaaagaccaaacaagcaltccagagcctgcagcagcaattctcaacaat 523
|||||
DB 421 aaagcaccacagaaagaccaaacaagcaltccagagcctgcagcagcaattctcaacaat 480
QY 524 gtcagctaagaagcttgcctcgtccttgtaagagctcgcagcagcagcagcagcagc 583
|||||
DB 481 gtcagctaagaagcttgcctcgtccttgtaagagctcgcagcagcagcagcagcagc 540
QY 584 aacattctcagcacaagaagacagctgagcacactaccagacactctctccaccctc 643
|||||
DB 541 aacattctcagcacaagaagacagctgagcacactaccagacactctctccaccctc 600
QY 644 actctccacatgtaccacacccctaaatcaltccagtgctctcaaaaagatgttttcaa 703
|||||
DB 601 actctccacatgtaccacacccctaaatcaltccagtgctctcaaaaagatgttttcaa 660
QY 704 gatcatcttctgtgtgctcctctctctctctctctctctctctctctctctctctct 763
|||||
DB 661 gatcatcttctgtgtgctcctctctctctctctctctctctctctctctctctctct 720
QY 764 ctccaccttaacagagcttagcttaattacacctgaagaattccaggaactgtagctctc 823
|||||
DB 721 ctccaccttaacagagcttagcttaattacacctgaagaattccaggaactgtagctctc 780
QY 824 agctagtcatcttaaccttaaatgcaatcagagaagtgcacaagaagtcaataata 883
|||||
DB 781 agctagtcatcttaaccttaaatgcaatcagagaagtgcacaagaagtcaataata 840
QY 884 ttttaataatgctc 895
|||||
DB 841 ttttaataatgctc 852
RESULT 7
AAS45994
ID AAS45994 standard; cDNA; 870 BP.
XX
XX AAS45994;
AC
XX
XX 18-DEC-2001 (first entry)
DT
XX
XX Human DNA encoding PRO polypeptide sequence #70.
DE
XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;

KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
 KW PCR primer.
 OS Homo sapiens.
 XX
 PN W0200168848-A2.
 XX
 XX 20-SEP-2001.
 PD
 XX
 PF 28-FEB-2001; 2001WO-US06520.
 XX
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 03-MAR-2000; 2000US-187202P.
 PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 14-MAR-2000; 2000US-189328P.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 29-MAR-2000; 2000US-193032P.
 PR 29-MAR-2000; 2000US-193053P.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 04-APR-2000; 2000US-194449P.
 PR 11-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 PA (GENTECH) GENENTECH INC.
 XX
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 XX WPI; 2001-602746/68.
 DR P-PSDB; AA029093.
 XX
 DR Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 XX Claim 2; Fig 139; 774pp; English.
 PS
 XX Sequences AAS45925-AAS4631 represent DNA molecules encoding and PCR
 CC primers for PRO polypeptides of the invention. The sequences of the
 CC invention can be used to detect the presence of a tumour in a mammal by
 CC comparing the level of expression of a PRO polypeptide in a test sample
 CC of cells from the animal and a control sample of normal cells, whereby a
 CC higher level of expression in the test sample indicates the presence of a
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,

CC pigs, goats and rabbits but are preferably human. The polypeptides can be
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human
 CC blood, when contacted with it. A specific polypeptide can be used to
 CC stimulate the proliferation or differentiation of chondrocyte cells. The
 CC PRO proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours. In mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 XX
 SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 22; Length 870;
 Best Local Similarity 100.0%; Pred. No. 6.3e-257;
 Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgcctcaaatgtggaacgtctgctggactaaagcatagaccacagctgagatc 103
 Db 1 ctgcgcctcaaatgtggaacgtctgctggactaaagcatagaccacagctgagatc 60
 QY 104 ctgacctgagtcacccccaaggatcaaggacccccaagggaaccttcattattct 163
 Db 61 ctgacctgagtcacccccaaggatcaaggacccccaagggaaccttcattattct 120
 QY 164 tcaagcaacttaagctgcgcacgaagtgcgtgaagttcaatctctccctctcc 223
 Db 121 tcaagcaacttaagctgcgcacgaagtgcgtgaagttcaatctctccctctcc 180
 QY 224 tgtgtcgtccactaatgtgctgctgctgctgctgctgctgctgctgctgctgct 283
 Db 181 tgtgtcgtccactaatgtgctgctgctgctgctgctgctgctgctgctgctgct 240
 QY 284 gaaggccaagggacgaaggccaagctctcagagatgctccagaagcgccgaagat 343
 Db 241 gaaggccaagggacgaaggccaagctctcagagatgctccagaagcgccgaagat 300
 QY 344 gtgagtcaaaagtgtgttcctcctgagagcccgccgaagaattcatgacagtgtctggc 403
 Db 301 gtgagtcaaaagtgtgttcctcctgagagcccgccgaagaattcatgacagtgtctggc 360
 QY 404 tgcgaagaagcagtgccctgtgcatcttcaaggaagcaattgaaagaacacacc 463
 Db 361 tgcgaagaagcagtgccctgtgcatcttcaaggaagcaattgaaagaacacacc 420
 QY 464 aaaggacacacgaagaagccaaacagcatccagagctgcagcaatttccaacaat 523
 Db 421 aaaggacacacgaagaagccaaacagcatccagagctgcagcaatttccaacaat 480
 QY 524 gtcaagctaagaagcttgcctgctgctgtgagagctctgagcgccacttccaatta 583
 Db 481 gtcaagctaagaagcttgcctgctgctgtgagagctctgagcgccacttccaatta 540
 QY 584 aacattctcagccaagaagaagcagtgagcacactaccagaacctctctccacctc 643
 Db 541 aacattctcagccaagaagaagcagtgagcacactaccagaacctctctccacctc 600
 QY 644 actctccactgtacccaccctaatatcatccagctgctctcaaaaagcagtgtttcaa 703
 Db 601 actctccactgtacccaccctaatatcatccagctgctctcaaaaagcagtgtttcaa 660
 QY 704 gatactttgttctgtctctctctagtgcttctctctctgctgaagcttgctgtgc 763
 Db 661 gatactttgttctgtctctctctagtgcttctctctctgctgaagcttgctgtgc 720
 QY 764 ctcccttaaccagagcttagtgtaattatcattcagaagctccagaactgtagctctc 823
 Db 721 ctcccttaaccagagcttagtgtaattatcattcagaagctccagaactgtagctctc 780
 QY 824 agctagtgatcatctaacttaaatgacatcaggaagtagcaagaagtagtaataa 883
 Db 781 agctagtgatcatctaacttaaatgacatcaggaagtagcaagaagtagtaataa 840

OY 884 ttcttaaatgctc 895
 |||
 Db 841 ttcttaaatgctc 852

RESULT 8
 AAF92070
 ID AAF92070 standard; cDNA; 870 BP.
 XX
 AC AAF92070;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Human PRO842 cDNA.
 XX
 KW Human; PRO protein; mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200116318-A2.
 XX
 PD 08-MAR-2001.
 XX
 PF 24-AUG-2000; 2000WO-US23328.
 XX
 PR 01-SEP-1999; 99WO-US20111.
 PR 15-SEP-1999; 99WO-US21090.
 PR 07-DEC-1999; 99US-0169495.
 PR 09-DEC-1999; 99US-0170262.
 PR 11-JAN-2000; 2000US-0175481.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 03-MAR-2000; 2000US-0187202.
 PR 25-APR-2000; 2000US-0199397.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 05-JUN-2000; 2000US-0209832.
 XX
 PA (GENE) GENENTECH INC.
 XX
 PI Eaton DL, Flivarooff E, Gerlitsen ME, Goddard A, Godowski PJ;
 PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
 XX
 DR MPI: 2001-183260/18.
 DR P-PSDB; AAB87538.
 XX
 PT Eighty four nucleic acids encoding PRO polypeptides, useful in
 PT molecular biology, including use as hybridization probes, and in
 PT chromosome and gene mapping. -
 XX
 PS Claim 2; Fig 25; 278bp; English.
 XX
 CC The present sequence is the coding sequence for a human PRO polypeptide
 CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO
 CC antagonists or anti-PRO antibodies are useful for preparation of a
 CC medicament useful in the treatment of a condition which is responsive to
 CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO
 CC protein may also be employed as molecular weight markers for protein
 CC electrophoresis. The PRO coding sequence has applications in molecular
 CC biology, including use as hybridisation probes, and in chromosome and
 CC gene mapping.
 XX
 SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 22; Length 870;
 Best Local Similarity 100.0%; Pred. No. 6.3e-257;
 Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 ctgcgcctcaaatggaacgctgctggactaagataagacacacagctgagtc 103
 |||
 Db 1 ctgcgcctcaaatggaacgctgctggactaagataagacacacagctgagtc 60

OY 104 ctgacctgagtcattcccccaggatcaaggagcctccagcaggagaccttcattatctc 163
 |||
 Db 61 ctgacctgagtcattcccccaggatcaaggagcctccagcaggagaccttcattatctc 120
 OY 164 tcaagcaacttaacagctgacccagcagctgacgtgaagaattcattctccctcc 223
 |||
 Db 121 tcaagcaacttaacagctgacccagcagctgacgtgaagaattcattctccctcc 180
 OY 224 tgttgcctcaactaaatgctgattgctcattgactagcagctgaaatccagggtcgca 283
 |||
 Db 181 tgttgcctcaactaaatgctgattgctcattgactagcagctgaaatccagggtcgca 240
 OY 284 gaggccacagggagccagcagcagcttctcaggatagcttccaggaagcggcagaagt 343
 |||
 Db 241 gaggccacagggagccagcagcagcttctcaggatagcttccaggaagcggcagaagt 300
 OY 344 gtgagtgcaaaagattgctcctgtagagcccgagaaatccatgacagtgctggc 403
 |||
 Db 301 gtgagtgcaaaagattgctcctgtagagcccgagaaatccatgacagtgctggc 360
 OY 404 tgcacaaagaagcagtgccctgtgatcatttcaaggcgaatgtgaaagaacagacacc 463
 |||
 Db 361 tgcacaaagaagcagtgccctgtgatcatttcaaggcgaatgtgaaagaacagacacc 420
 OY 464 aaagcaccacacagaaagcacaacacagcattccagagccctgcagacatttccaaacat 523
 |||
 Db 421 aaagcaccacacagaaagcacaacacagcattccagagccctgcagacatttccaaacat 480
 OY 524 gtcaagctaaagaagcttgctcgtccttgtagagctcgtgagcgccactctccaat 583
 |||
 Db 481 gtcaagctaaagaagcttgctcgtccttgtagagctcgtgagcgccactctccaat 540
 OY 584 aacatctcagcacaagaagacagtgagcagacactacagacactcttctccaccc 643
 |||
 Db 541 aacatctcagcacaagaagacagtgagcagacactacagacactcttctccaccc 600
 OY 644 actctccacctgtaacccacccctaaatcattcagagctcctcaaaagatgttttcaa 703
 |||
 Db 601 actctccacctgtaacccacccctaaatcattcagagctcctcaaaagatgttttcaa 660
 OY 704 gatcatttgttctgtcctcctcctagtgctctctctctctctcgtcactagcctgac 763
 |||
 Db 661 gatcatttgttctgtcctcctcctagtgctctctctctctcgtcactagcctgac 720
 OY 764 ctcccttaacccaagcgttaagcttaactacactgaagaattccaggaactgtagctcct 823
 |||
 Db 721 ctcccttaacccaagcgttaagcttaactacactgaagaattccaggaactgtagctcct 780
 OY 824 agctagtgctcatttaaccttaaatgtaacaggaagtagcaaaagagctcaataata 883
 |||
 Db 781 agctagtgctcatttaaccttaaatgtaacaggaagtagcaaaagagctcaataata 840
 OY 884 ttcttaaatgctc 895
 |||
 Db 841 ttcttaaatgctc 852

RESULT 9
 AAF44147
 ID AAF44147 standard; cDNA; 870 BP.
 XX
 AC AAF44147;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Human PRO842 (UNQ473) nucleotide sequence SEQ ID NO:164.
 XX
 KW Human; secreted and transmembrane protein; PRO; cytosolic;
 KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
 KW diagnostic assay; ss.
 XX
 OS Homo sapiens.

XX WO200073454-A1.
 XX 07-DEC-2000.
 XX 30-MAR-2000; 2000WO-US08439.
 XX 02-JUN-1999; 99WO-US12252.
 PR 23-JUN-1999; 99US-0141037.
 PR 07-JUL-1999; 99US-0143048.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 28-JUL-1999; 99US-0146222.
 PR 17-AUG-1999; 99US-0149396.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 08-OCT-1999; 99US-0158663.
 PR 30-NOV-1999; 99WO-US28313.
 PR 01-DEC-1999; 99WO-US28301.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00376.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 20-MAR-2000; 2000WO-US07377.
 XX
 PA (GERTH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Bolstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
 PI Grimaldi CJ, Gurney AL, Kijavini IU, Napier MA, Pan J, Paoni NF;
 PI Roy MA, Stewart TA, Tunas D, Watanabe CK, Williams PM, Wood WI;
 PI Zhang Z;
 XX
 DR WPI: 2001-032160/04.
 DR P-PSDB: AAB65191.
 XX
 PT PRO polynucleotides used to produce polypeptides used to target
 PT bioactive molecules such as toxins, radiolabels or antibodies, to
 PT specific cells, to cause targeted cell death.
 PT
 XX
 PS Claim 2; Fig 98; 935pp; English.
 PS
 XX
 CC The present invention describes human secreted and transmembrane PRO
 CC proteins. The PRO proteins have cytosolic activity. The PRO proteins
 CC can be used for targeted delivery of bioactive molecules, such as
 CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
 CC sequences, and their fragments, can be used as hybridisation probes, in
 CC chromosomal and gene mapping, and in the generation of anti-sense RNA
 CC and DNA. They may also be used to produce transgenic animals which are
 CC used to develop and screen therapeutically useful reagents. The PRO
 CC nucleotide and protein sequence can be used for tissue typing and in
 CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
 CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
 CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
 CC AAB65134 to AAB65300 represent human PRO polynucleotide and protein
 CC sequences given in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 22; Length 870;
 Best Local Similarity 100.0%; Pred. No. 6.3e-257;
 Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgcctaaatgggaagcgtgctggaactaagacacacagcgtggtatc 103
 |||
 |||

Db 1 ctgcgcctaaatgggaagcgtgctggaactaagacacacagcgtggtatc 60
 QY 104 ctgacctgagtcaccccaaggatcagagagctcccaagcaggaacttcattatct 163
 Db 61 ctgacctgagtcaccccaaggatcagagagctcccaagcaggaacttcattatct 120
 QY 164 tcaagcaacttaagcgtgcacccagcaatctgcataagaaagtcttactctccctcc 223
 Db 121 tcaagcaacttaagcgtgcacccagcaatctgcataagaaagtcttactctccctcc 180
 QY 224 tgttctgccaactaatgtcatgtccatgtgtctcctaagcctgaatccaggggtccca 283
 Db 181 tgttctgccaactaatgtcatgtgtctcctaagcctgaatccaggggtccca 240
 QY 284 gaggccacagggacagcagcagcagcgtctcagagatgtgtccaggaagcgcagaat 343
 Db 241 gaggccacagggacagcagcagcagcgtctcagagatgtgtccaggaagcgcagaat 300
 QY 344 gtgagtgcaagatgtgtctcctaagagcccccaggaagaatctcagacagtgctgggc 403
 Db 301 gtgagtgcaagatgtgtctcctaagagcccccaggaagaatctcagacagtgctgggc 360
 QY 404 tgcgaagaagcagtgccctgtgacatctcagaagcgaatgtgaagaacaagacacc 463
 Db 361 tgcgaagaagcagtgccctgtgacatctcagaagcgaatgtgaagaacaagacacc 420
 QY 464 aaaggacacacagaaagcacaacacatctccagagcgtccagcaattctcaacaat 523
 Db 421 aaaggacacacagaaagcacaacacatctccagagcgtccagcaattctcaacaat 480
 QY 524 gtcaagtaagaagcttgctctgtccttctgtagagctcgaagccactcttcaata 583
 Db 481 gtcaagtaagaagcttgctctgtccttctgtagagctcgaagccactcttcaata 540
 QY 584 aacattctcagcacaagaagcagtgagacacactaccagacactcttctccaccc 643
 Db 541 aacattctcagcacaagaagcagtgagacacactaccagacactcttctccaccc 600
 QY 644 actctccactgtaccacccccaataatcttcagagctctcaaaaagatgttttcaa 703
 Db 601 actctccactgtaccacccccaataatcttcagagctctcaaaaagatgttttcaa 660
 QY 704 gatcaattgtttgtctctctcctagtgctctctctcctcagcttaagcctgtgccc 763
 Db 661 gatcaattgtttgtctctctcctagtgctctctctcctcagcttaagcctgtgccc 720
 QY 764 ctcccttaaccaggttaggttaattaccctggaagatctcaggaagaactgtagctctc 823
 Db 721 ctcccttaaccaggttaggttaattaccctggaagatctcaggaagaactgtagctctc 780
 QY 824 agctagtgcatcttaaccttaataatgcaatcagaagaagtagcaaacagatcaataa 883
 Db 781 agctagtgcatcttaaccttaataatgcaatcagaagaagtagcaaacagatcaataa 840
 QY 884 ttlttaaatgctc 895
 Db 841 ttlttaaatgctc 852

RESULT 10
 AAA08343
 ID AAA08343 standard; DNA; 849 BP.
 XX
 AC AAA08343;
 XX
 DT 30-JUN-2000 (first entry)
 DE Human TGC-440 secretory protein nucleotide sequence.
 XX
 KW TGC-440; secretory protein; immunological disease; infectious disease;
 KW pulmonary function disorder; hepatic function disorder; nephrotropic;
 KW gastrointestinal function disorder; antiinflammatory; immunomodulatory;
 KW virucide; hepatotropic; antisthmatic; antibacterial; vaccine;

CC inhibition and in vaccines. The NAS and the lung tumour-associated
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with their inappropriate expression,
CC especially lung cancers. For example, the NAS may be administered to
CC treat diseases by rectifying mutations or deletions in a patient's genome
CC that affect the activity of the protein by expressing inactive proteins
CC or to supplement the patient's own production of (1). Additionally, the
CC NAS may be used to produce the lung-tumour associated protein, according
CC to standard recombinant DNA methodology. Conversely, antisense NA
CC molecules may be administered to down regulate protein expression by
CC binding with the cells own genes and preventing their expression. The NA
CC and complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar NA sequences in
CC samples, and hence which patients may be in need of treatment for lung
CC cancer. The (1) may be used as antigens in the production of antibodies
CC and in assays to identify modulators (agonists and antagonists) of the
CC expression and activity of the protein. AAF68083 to AAF6878 and
CC AAF6848 to AAF6878 represent human lung tumour protein related
CC nucleotide and protein sequences which are used in the exemplification
CC of the present invention.

CC
XX
S0 Sequence 533 BP; 152 A; 154 C; 113 G; 114 T; 0 other:

Query Match 58.7%; Score 525.8; DB 22; Length 533;
Best Local Similarity 99.6%; Pred. No. 1.3e-154;
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 179 ctgacgcagcagtgatgataaagttctaattcttcctcctcctctgtgtgtccactaa 238
Db 1 ctgcacgcagcagtgatgataaagttctaattcttcctcctcctctgtgtgtccactaa 60
OY 239 tgcgtatgtccatgtgtctctctgacgacctgaatccaggggtgcgcagagccacaggacc 298
Db 61 tgcgtatgtccatgtgtctctctgacgacctgaatccaggggtgcgcagagccacaggacc 120
OY 299 gaggcacagcgtctcagagatggtccacagagcgcccaagaatgtagtcaagaatt 358
Db 121 gaggcacagcgtctcagagatggtccacagagcgcccaagaatgtagtcaagaatt 180
OY 359 ggttcctgagagcccgagaaataatcattgacagtgctctgtggtccaaagaagcagt 418
Db 181 ggttcctgagagcccgagaaataatcattgacagtgctctgtggtccaaagaagcagt 240
OY 419 gccctgtgattcattcaaggcgaatgtgaaagaaacaaagacacacagacacagaa 478
Db 241 gccctgtgattcattcaaggcgaatgtgaaagaaacaaagacacacagacacagaa 300
OY 479 agccaaacaaagcattccagagcctgcagcaattctcaacaatgtcagctaaagaagt 538
Db 301 agccaaacaaagcattccagagcctgcagcaattctcaacaatgtcagctaaagaagt 360
OY 539 ttgtctgtcctttagtagagctctgagcgcccaactcttcaacatgtaacattccagcaa 598
Db 361 ttgtctgtcctttagtagagctctgagcgcccaactcttcaacatgtaacattccagcaa 420
OY 599 gaagcagatgagcacactaccagacactcttctctccacactcaactctccactgtac 658
Db 421 gaagcagatgagcacactaccagacactcttctctccacactcaactctccactgtac 480
OY 659 ccacccccaatcattccagtgctctcaaaaagcattgtttccaagatc 707
Db 481 ccacccccaatcattccagtgctctcaaaaagcattgtttccaagatc 529

RESULT 13
AAV69618
ID AAV69618 standard; DNA: 511 BP.

XX AC AAV69618;
XX DT 28-JAN-1999 (first entry)
XX

DE Human secreted protein gene 8 clone HLHCM89.

XX Secreted protein; gene therapy; protein therapy; diagnosis; treatment;
KW central nervous system; CNS; immune system; cancer; trauma; liver;
KW reproductive disorder; congenital malformation; degenerative disease;
KW inflammatory disease; neoplasia; metabolic disorder; testis; placenta;
KW brain; T cell; spleen; lung; heart; rhabdomyosarcoma; endocrine system;
KW endocrinopathy; endocrine polyglandular syndrome; endocrinoma; sepsis;
KW endocrine ophthalmopathy; osteoclastoma; bacterial infection; bone; ds.

OS Homo sapiens.

XX Key Location/Qualifiers
XX CDS 233..427

FT /tag= a
FT /product= "secreted protein"
FT /transl_except= (pos:344..346, aa:Xaa)
FT /transl_except= (pos:383..385, aa:Xaa)
FT /note= "Xaa = unknown"

FT sig_peptide 233..310
FT /tag= b
FT mat_peptide 311..424
FT /tag= c

W09845712-A2.

15-OCT-1998.

07-APR-1998; 98WO-US06801.

XX 30-MAY-1997; 97US-0048184.
XX 08-APR-1997; 97US-0042726.
XX 08-APR-1997; 97US-0042727.
XX 08-APR-1997; 97US-0042728.
XX 08-APR-1997; 97US-0042754.
XX 08-APR-1997; 97US-0042825.
XX 30-MAY-1997; 97US-0048068.
XX 30-MAY-1997; 97US-0048070.

(HUMA-) HUMAN GENOME SCT INC.

Feng P, NI J, Rosen CA, Ruben SM, Yu G;

WPI: 1998-594496/50.

P-PSDB; AAM83938.

PT New isolated human genes and secreted polypeptide(s) they encode -
PT useful for the diagnosis and treatment of e.g. cancers, CNS
PT disorders, immune system disorders, inflammatory disease and
PT bacterial infections

PS Claim 4; Page 107; 142pp; English.

XX This sequence represents a nucleic acid molecule designated Gene 8 from
XX the human CDNA clone HLHCM89 (deposited as clone ATCC 97955 and ATCC
XX 209074) which encodes a secreted human protein. This gene is expressed
XX primarily in lung and to a lesser extent in pancreatic carcinoma and
XX gall bladder and is useful as reagents for differential identification
XX of tissues in a biological sample.

CC The invention relates to 20 novel genes and their fragments (AAV69611 to
CC AAV69630) and corresponding secreted proteins (AAM83931 to AAM83950)

CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein of gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the
CC polynucleotides. Specific uses are based on which tissues they are most
CC highly expressed in, and include developing products for the diagnosis or
CC treatment of central nervous system (CNS) and immune system diseases,
CC reproductive disorders, cancers, congenital malformations, degenerative
CC diseases, trauma, inflammatory disease, neoplasia, metabolic disorders,
CC diseases in testes, placenta, liver, brain and activated T cells, spleen
CC diseases, lung diseases, heart diseases, rhabdomyosarcoma and disorders
CC of the endocrine system or other endocrinopathies, e.g. endocrine

CC polyglanular syndrome, endocrinoma, and endocrine ophthalmopathy,
CC osteoclastoma and other bone remodelling disorders, bacterial infections
CC and sepsis. The polypeptides are also useful for identifying their
CC binding partners.
XX

SQ Sequence 511 BP; 138 A; 138 C; 125 G; 104 T; 6 other;

Query Match 49.7%; Score 445.2; DB 19; Length 511;
Best Local Similarity 97.9%; Pred. No. 2,6e-129;
Matches 466; Conservative 4; Mismatches 4; Indels 2; Gaps 2;

```
OY 1 ctaatctgttaagtaacgaagacagcgtccactcctgttctccgccaataggga 60
DB 38 ctaatctgttaagtaacgaagacagcgtccactcctgttctccgccaataggga 97
OY 61 acgctgcttggaactaagacacacagcgtgactcctgactgactgactgactc 120
DB 98 acgctgcttggaactaagacacacagcgtgactcctgactgactgactgactc 157
OY 121 cagggatcagagcctccagcaggaaccttcattatattctcaagaactaacag 180
DB 158 cagggatcagagcctccagcaggaaccttcattatattctcaagaactaacag 217
OY 181 gacacagacgttgatgaagtttaattctctccctccctcctgctgctgactaatg 240
DB 218 gacacagacgttgatgaagtttaattctctccctccctcctgctgctgactaatg 277
OY 241 ctgattcattatgctctcagcagcctgactcagaggtgcagcagacagagacagga 300
DB 278 ctgattcattatgctctcagcagcctgactcagaggtgcagcagacagagacagga 337
OY 301 gacacagccttctagagatgctcctcaggaagcggcgaagaattgagtgcaaatgg 360
DB 338 gacacagccttctagagatgctcctcaggaagcggcgaagaattgagtgcaaatgg 397
OY 361 ttctctgaagcggcgaagaagaattctcagcagtgctgctgctgctgctgctgctg 420
DB 398 ttctctgaagcggcgaagaagaattctcagcagtgctgctgctgctgctgctgctg 456
OY 421 cccgtgtgcttctcagagggcagatgcaagaacacacacacacacacacacag 476
DB 457 cccgtgtgcttctcagagggcagatgcaagaacacacacacacacacacacag 511
```

RESULT 14
AAx40454
ID AAx40454 standard; cDNA; 365 BP.
XX
AC AAx40454;

DT 18-JUN-1999 (first entry)

XX Human secreted protein 5' EST SEQ ID No: 54.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokine; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.

OS Homo sapiens.
XX
XX
PN MO9906550-A2.
XX
PD 11-FEB-1999.
XX
XX
PF 31-JUL-1998; 98WO-IB01232.
XX
PR 01-AUG-1997; 97US-0905144.
XX
PA (GSEST) GENSET.

XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX WPI; 1999-153780/13.
DR P-PSDB; AAY11732.
XX

PT New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
PS Claim 1; Page 185; 675pp; English.

XX AAx40438 to AAx40715 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins expressed in prostate, and encode the proteins
CC given in AAY11716 to AAY11993 respectively. The proteins given represent
CC the signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation and differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokine activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptides can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX

SQ Sequence 365 BP; 91 A; 100 C; 96 G; 78 T; 0 other;

Query Match 40.5%; Score 362.4; DB 20; Length 365;
Best Local Similarity 99.7%; Pred. No. 2.1e-103;
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY 39 ctgttcgcccctcaaatggaacgtgctggcactcaagcatagacacacagctga 98
DB 2 ctgttcgcccctcaaatggaacgtgctggcactcaagcatagacacacagctga 61
OY 99 gttactcgtgactgactcctccagggatcagagacctccagcagagacattccat 158
DB 62 gttactcgtgactgactcctccagggatcagagacctccagcagagacattccat 121
OY 159 attcttcaagcaacttaagcgtgcacgcagcgttgctgctgaagtctcattccct 218
DB 122 attcttcaagcaacttaagcgtgcacgcagcgttgctgctgaagtctcattccct 181
OY 219 cctctgtgtgctgcaactatgctatgctcctcctcagcagcgtgaattccagggt 278
DB 182 cctctgtgtgctgcaactatgctatgctcctcctcagcagcgtgaattccagggt 241
OY 279 cgcagagggcagcagggagcagcagcgtctcagagatgctcagagagcgagcca 338
DB 242 cgcagagggcagcagggagcagcagcgtctcagagatgctcagagagcgagcca 301
OY 339 agaattgtagcgaagtgctgctcagagagcggcgaagaagaattcattgacagtgc 398
DB 302 agaattgtagcgaagtgctgctcagagagcggcgaagaagaattcattgacagtgc 361
OY 399 tggg 402
DB 362 tggg 365
```

RESULT 15
AA08344
ID AA08344 standard; DNA; 357 BP.
XX
AC AA08344;
XX
DT 30-JUN-2000 (first entry)

| XX | Human TGC-440 secretory protein nucleotide sequence SFO ID NO:4. |
|----|--|
| DE | |
| XX | |
| KW | TGC-440; secretory protein; immunological disease; infectious disease; |
| KW | pulmonary function disorder; hepatic function disorder; nephrotropic; |
| KW | gastrointestinal function disorder; antiinflammatory; immunomodulatory |
| KW | virulence; hepatotropic; antiasthmatic; antibacterial; vaccine; |
| KW | hepatitis; nephritis; influenza; asthma; pulmonary hypertension; |
| KW | pneumonia; Helicobacter pylori infection; ds. |

| | |
|----|---|
| OS | Homo sapiens. |
| XX | |
| FH | Key |
| FT | CDS |
| FT | Location/Qualifiers |
| FT | 1..357 |
| FT | /*tag= a |
| FT | /product= "TGC-440" |
| FT | /note= "secretory protein; no stop codon given" |

WO200014226-A1.

16-MAR-2000.

PF 02-SEP-1999; 99WO-JP04765.

PR 03-SEP-1998; 98JP-0250108.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Itoh Y, Ogi K, Tanaka H,

DR WPI; 2000-256978/22.

XX

PT inhibiting its activity for diagnosis and treatment of diseases of the

PT immune system, lung, kidney, liver and intestinal system -

PS Disclosure; Page 79; 86pp; Japanese

3 XX

CC TGC-440 has antiinflammatory, nephrotropic, immunomodulatory,

CC and can be used in vaccines. TGC-440 and the polynucleotide sequence

lung, liver, kidney or gastrointestinal disorders and infectious

CC pulmonary hypertension, and *Helicobacter pylori* infection. An antibody

diagnosis, and also for quantifying the amount of TGC-440 in a liquid

XXXXXXXXXXXX

Sequence 357 BP; 102 A; 92 C; 90 G; 73 T; 0 other;

| | | | | |
|-------------|--------|------------|--------|-------------|
| Query Match | 39.98; | Score 357; | DB 21; | Length 357; |
|-------------|--------|------------|--------|-------------|

| | | | | | | | | | |
|---------|------|--------------|----|------------|----|--------|----|------|----|
| Matches | 357; | Conservative | 0; | Mismatches | 0; | Indels | 0; | Gaps | 0; |
|---------|------|--------------|----|------------|----|--------|----|------|----|

196 atgaagttctaattccctcctgtgtgctgccataatgctgatgtccatgctc 25

Db 1 atgaagttctaattcctctccctgtgtgctgcaataatgctgatgtc 60

QY 256 tctagcagcctgaatccagggtcgcagagccacagggaccgagccagctctag 31

Db 61 tctagcagccctgaatccaggggtcgcagagggccaagggacgcagggcctctag 12

316 agatggtccaggaagggccaagatgtgagtcaagattgttctgagagcccg 37

db 121 agatggtccaggaagcgccaagaattgtgagtgcaagaattgttcttgagagcccg 18

3/b agdaagaaattcatgacagtgtctggctgccaagaagcagtgccctgtgatcatctc 43

DB 181 agaaagaaaattcattgacagtgctctggtccaaagaagcagtgccctgtgatcatctc 24

QY 436 aagggcaatgtgaagaaaacaagacacccaagggcaccacagaanaagccaacaagcatcc 495

Db 241 aagygcatgtgaagaaacaagacaccaaagcaccacagaagaagccaacaagcatcc 300

QY 496 agagcctgcagcaatttctcaaacatgtcagctaagaagcttgcctctgccttg 552

Db 301 agagcctgcagcaatttctcaacaatgtcagctaagaagcttgcctctgctccttg 357

Search completed: September 16, 2002, 03:05:13
Job time: 7991 sec

| | | | |
|----|-----|---|-----|
| Oy | 436 | aaggcaatgtgaaagaaaacagcaccacaagggccacgcagaagaaagccaacaagcatccc | 495 |
| Db | 241 | aaggccaatgtgaaagaaaacagcaccacaagggccacgcagaagaaagccaacaagcatccc | 300 |
| Oy | 496 | agagctcgccagcaattctctaaacaatgttcagttaagaagcttgcctgcgcttg | 552 |
| Db | 301 | agagctcgccagcaattctctaaacaatgttcagttaagaagcttgcctgcgcttg | 357 |

...



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: September 16, 2002, 00:47:47 ; Search time 3013.46 Seconds
(without alignments)
7524.535 Million cell updates/sec

Title: US-09-700-770-2
Perfect score: 1680
Sequence: 1 ggtgtgcagataaagt.....tcaataaacactgtctgtg 1680

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapect 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: qb_est1:*
10: qb_est2:*
11: qb_hlc:*
12: qb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 835.2 | 49.7 | 945 | 10 | BI772722 603053235 |
| 2 | 791.2 | 47.1 | 852 | 10 | BI818205 603032679 |
| 3 | 771 | 45.9 | 783 | 10 | BI770190 603053221 |
| 4 | 762.4 | 44.5 | 807 | 10 | BI823956 603039139 |
| 5 | 746.8 | 44.4 | 784 | 10 | BI821790 603035865 |
| 6 | 731.8 | 43.6 | 826 | 10 | BG529820 602558904 |
| 7 | 702.6 | 41.8 | 785 | 10 | BG530180 602558678 |
| 8 | 684.6 | 40.8 | 955 | 10 | BI821769 603035838 |
| 9 | 671.6 | 40.0 | 836 | 10 | BI820886 603033924 |
| 10 | 624.2 | 37.2 | 660 | 10 | BG548906 602575739 |
| 11 | 623.2 | 37.1 | 796 | 10 | BG540252 602569143 |
| 12 | 619.2 | 36.9 | 912 | 10 | BG484717 602505896 |
| 13 | 618 | 36.8 | 658 | 9 | BE061388 OVO-BT022 |
| 14 | 604 | 36.0 | 642 | 10 | BF672220 602150403 |
| 15 | 598.8 | 35.6 | 640 | 10 | BG482968 602502981 |
| 16 | 588.8 | 35.0 | 754 | 10 | BG547238 602574650 |
| 17 | 587 | 34.9 | 602 | 10 | BG530024 602558870 |

| | | | | | |
|----|-------|------|------|----|---------------------|
| 18 | 581.4 | 34.6 | 630 | 10 | BG482764 602502549 |
| 19 | 559.2 | 33.3 | 743 | 10 | BG548174 602575330 |
| 20 | 537 | 32.0 | 549 | 10 | BG482718 602502690 |
| 21 | 521.4 | 31.0 | 539 | 10 | BI824243 603040625 |
| 22 | 519.6 | 30.9 | 567 | 10 | BF088537 RC1-HT088 |
| 23 | 519.4 | 30.9 | 556 | 10 | BE933205 RC1-HT088 |
| 24 | 513.8 | 30.6 | 607 | 10 | BE720102 RC1-HT088 |
| 25 | 511 | 30.4 | 525 | 9 | AL602621 DKFZP686K |
| 26 | 484.6 | 28.8 | 524 | 10 | BE933204 RC1-HT088 |
| 27 | 473.4 | 28.2 | 541 | 9 | AM384404 CM2-HT039 |
| 28 | 463.2 | 27.6 | 469 | 9 | A1924632 wnt7e12.x |
| 29 | 443.8 | 26.4 | 482 | 10 | BF095122 PM3-UT005 |
| 30 | 439.8 | 26.2 | 465 | 10 | BF154912 RC0-BT081 |
| 31 | 434.8 | 25.9 | 438 | 9 | AM167650 xnat7a11.x |
| 32 | 430.6 | 25.6 | 463 | 10 | BF743981 RC0-BT081 |
| 33 | 426.4 | 25.4 | 1026 | 10 | BI821994 RC0-BT081 |
| 34 | 419.4 | 25.0 | 957 | 10 | BI822130 RC0-BT081 |
| 35 | 417.8 | 24.9 | 421 | 10 | BF154920 RC0-BT081 |
| 36 | 404.8 | 24.1 | 433 | 10 | BF826403 CM4-HN002 |
| 37 | 404 | 24.0 | 428 | 10 | BE932793 CM2-HT087 |
| 38 | 401.8 | 23.9 | 429 | 9 | AM449208 UT-H-BT3- |
| 39 | 392.2 | 23.3 | 397 | 10 | BE693858 RC3-BT055 |
| 40 | 389.4 | 23.2 | 576 | 9 | AM363590 CM1-CT033 |
| 41 | 380.4 | 22.6 | 402 | 10 | BF831058 CM1-HT087 |
| 42 | 373.4 | 22.2 | 390 | 10 | BE829540 RC0-ET007 |
| 43 | 361.4 | 21.5 | 363 | 9 | AM167610 xnat9e05.x |
| 44 | 354.4 | 21.1 | 379 | 10 | BF088529 RC1-HT088 |
| 45 | 353.6 | 21.0 | 360 | 9 | BE073612 RC5-BT055 |

ALIGNMENTS

RESULT 1
LOCUS BI772722 945 bp mRNA linear EST 25-SEP-2001
DEFINITION 603053235P1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5202801 5',
mRNA sequence.
ACCESSION BI772722
VERSION BI772722.1 GI:15764300
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE 1 (bases 1 to 945)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LHAM11508 row: e column: 10
High quality sequence stop: 856.
Location/Qualifiers
1..945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5202801"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen. Vector: pCMV-SPORT6;
Site.1: NotI, Site.2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size

FEATURES
source


```

|||||
Db 241 CTGCTCAGTGCATGGGAGAAAGCCAGGAGCATCCCTGCTGGAGACCTGTG 300
|||
QY 354 aacacgcgtctgaagaacatctatctgctgaaggtatcatcaagctaaatctcagctg 413
|||
Db 301 AACACCGTCTTAAGACATCATCTGGCAAGGTCATACAGCTAACATCTCCACGCTG 360
|||
QY 414 caggtgaacccctcgcccatgaccagagctgctgaatgaatccccctggaacatggtg 473
|||
Db 361 CAGGTGAACCCCTCGGCCAATGACAGAGCTGCTAGTAAGATCCCCCTGGACATGTG 420
|||
QY 474 gctgattcaaacagccccctggtcaagaacatcgtggaattccacaatgaatgaatgagcc 533
|||
Db 421 GCTGATTCAACACGCCCCCTGGTCAAGACCATGTTGAGATTCACATGACGATGAGGCC 480
|||
QY 534 caagccacatctcgatgtagacacccagtgaaagtgtgccccacccgctgtctcagtgac 593
|||
Db 481 CAAAGCCACCATCCGATGGACACCAAGTGCAGTGGCCACCCGCTGGTCTCAAGTGAC 540
|||
QY 594 tgtgcacacagccatgtagagctgctgcacacactgctgataagctctcctcctgtg 653
|||
Db 541 TGTGGCACACGACCATGGAGGCTGGGACATCAACTGCTCATAGCTCTCTTCCTGGTG 600
|||
QY 654 aacgccttagctcaagcaggtcatgaactcctcctagtgcaatccctgccaatctagtga 713
|||
Db 601 AACGCTTAGCTAAGACGCTCATGAACCTCTAAGTGCATCCCTGCCAATCATAGTAA 660
|||
QY 714 aacccagctgtgtccctgtagtgaggtcttcaatgagatgatgaacatccctgca 772
|||
Db 661 AACCCAGCTGTGTCCCTGTATGCAAGGCTTCTTCAATGAGCATGTCACAACTCCCTGCA 720
|||
QY 773 gctgtgaaggtgtccatctcctcagcatgacgctcgg-agttagacctctgtac 831
|||
Db 721 GCTGTGAAGGTGTCCATTTCCCTCAGCATGACCTCTGGAAGTGTGACTTCTGTATC 780
|||
QY 832 ctgcacatgaaggtgacacacatcagctcactcgg-aggccaagttgttgactacag 890
|||
Db 781 CTGCGCATCAAGGGGTGACACCATTCAGCTTACTGAGGGGCAAGTTGGGAGCTACCG 840
|||
QY 891 gga 894
|||
Db 841 gga 844

```

```

RESULT 3
LOCUS B1770190 783 bp mRNA linear EST 25-SEP-2001
DEFINITION 603053221F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5202893 5',
mRNA sequence.
ACCESSION B1770190
VERSION B1770190.1 GI:15761768
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
Unpublished (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1M11508 row: 1 column: 06
High quality sequence stop: 782.
FEATURES
source 1..783

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5202893"
/clone_1ib="NIH_MGC_122"
/lab_host="DH10B"
/note="organ: pooled lung and spleen; Vector: pCMV-Sport6;
site_1: Not; Site-2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
BASE COUNT 167 a 261 c 199 g 156 t
ORIGIN

```

```

Query Match 45.9%; Score 771; DB 10; Length 783;
Best Local Similarity 99.9%; Pred. No. 7.4e-176;
Matches 782; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

```

QY 53 ggaagcgagcgagagcagcagcgtgcccaggtctgcatcctgcaactgtgctcctcga 112
|||
Db 1 GGAGCGGGCGGAGAGACTCCAGGCTGCCAGTGTGACATCTGCACTTGCTGCTCTGA 60
|||
QY 113 cacttggaagatgagcgcccggtggaactcaccctctctgtgtgttctgagcgca 172
|||
Db 61 CACCTGGGAAGATGGCGGCCCTGGACCTTCACCTCTCTGTGGTTGGTCGACGCA 120
|||
QY 173 ccttgatcaagcaccctccagtcaccagtcagttcactcctcctgcccagaatcata 232
|||
Db 121 CCTTGATCCAGCCACCTTCAGTCCACATGCACTTCTCATCTCTCGGCCCAAAAGTATGA 180
|||
QY 233 aagaaagctgacacagagagctgaagagaccacagccacacagcctcctcagcagctgc 292
|||
Db 181 AAGAAAGCTGACACACAGAGCTGAAGAGACACAAACCCACACAGCATCTGACAGACTGC 240
|||
QY 293 cgtctcagtgatcagatgagcggaagaaagcagcggaag-attcctgtgctggcgagctg 351
|||
Db 241 CGCTGCTCAAGTGCATGCGGGAAGAACCGGAGGATCCCTGTGCTGGGAGGCTGTG 300
|||
QY 352 tgaacacgctcctgtaagacatcatctgtgctgaaggtcatcaagctaacatcctcagc 411
|||
Db 301 TGAACACCGTCTTGAGACACATCATCTGGCTGAAGGTCTACAGCTTAACATCTCTCAGC 360
|||
QY 412 tgaaggtgaagcctcctgacaaatgacagagagctgtgtaagaatccccctggaatg 471
|||
Db 361 TGCAGGTGAAGCCCTCGGCCAATGACCAAGAGAGCTGTAGCAAGATCCCTGTGACATG 420
|||
QY 472 tggctgattcaaacagccccctgtgtcaagaacatcgtgtgaagtctcaatgaagactgag 531
|||
Db 421 TGGCTGTGATTCAACACGCCCCCTGGTCAAGACATCTGTGAGTCCCATGACACGTAGG 480
|||
QY 532 cccaagcacaacatcgcatgtagaacacagtgcaagtgtgcccacccgctggtcctcagtg 591
|||
Db 481 CCCAAGCCACATCCGATGAGACACCAAGTGAAGTGGCCACCCGCTGTGCTCAGTG 540
|||
QY 592 actgtgccaacagcctggaagcctgctgcatcacaactgtgcaataagctctcctcctg 651
|||
Db 541 ACTGTGCACACGACCATGGAGGCTCGCATCAACTGTGCTGCTGCTGCTGCTGCTGCTG 600
|||
QY 652 tgaagccttagtgaagcaggtgtcaactcctcctgcaatccctgcccacatctagta 711
|||
Db 601 TGAACCCCTTAGCTAAGCAGGTATGAACCTCTCTGATGCTGCTGCTGCTGCTGCTG 660
|||
QY 712 aaaaacagctgtgtccggtgatcgaagctcctcctaagtgatgtatgaagactcctgc 771
|||
Db 661 AAAACCAAGCTGTCTCCGATGATGAGGCTTCTTCAATGACATGATGACAGACTCTG 720
|||
QY 772 agctgtgaaggtgccaatctcctcagcatatgacagctgtggaattgacactctgtac 831
|||

```



```

OY 222 aaaaatcatcaaaagaagctgacacagagctgaagaccacaagccacagatcct 281
Db 182 AAAAGTCATCAAGAAAGAGCTGACACAGAGAGCTGAAGAGCCACACAGCCACAGATCCT 241
OY 282 gcaagagctgacgctgctcactgacatgacatgacgagaaagcagccgagag-attcctgct 340
Db 242 GCAGCAGCTGCGCTCTCAGTCCATGCGGGAAAAAGCCAGCGGAGCATCCCTGTGCT 301
OY 341 gggcagcctggtgaacacgctcctgaagacatcctcgtgagagatgatacagctaa 400
Db 302 GGGCAGCCTGCTGAACACCGTCTCTGAGACATCATCTTGCTGAAGGTATCATCAGACTAA 361
OY 401 catctccagctgacgagtgtaagccctcggcacaatgacacagagagctgctgaatgaatccc 460
Db 362 CATCTCCAGCTGACAGGTGAAGCCCTCGGCAATGACCGAGAGCTGCTAGTAAATATCCC 421
OY 461 cctggaacatggtgctgagatcaacagccctcgtgtaagacacatcgtggaattccacat 520
Db 422 CCTGGACATGCTGCTGATTAACAACGCCCTGCTGAAGACCATCTGGAAGTTCACAT 481
OY 521 gacgaactgagcccaagccacatccgcatggaacacagtgcaatgagccacccgct 580
Db 482 GACGACTGAGGCCAAGCCACCATCCGATGAGACACAGTGAAGTGGCCCGCCGCT 541
OY 581 ggtcctcagtgactgtgacacagcagtgagagcctgcatccaactgtgtaagct 640
Db 542 GGTCTCTAGTACTGTGCCACCGCATGGAGCCTGCGCATCCAACTGCTGCTAAAGCT 601
OY 641 ctctctcctggtgaagcctt-agctaagcagtgatcaatacctcctgctgacatccctgc 699
Db 602 CTCTCTCTGTGTAAGCCTTAAGCTAGAGGTGATCAACCTCTCTAGTGCATCCCTGC 661
OY 700 ccaatcctgtaaaaaaacagctggtccc--gtatcgaggtcctcctcaatgacatgta 757
Db 662 CCAATCTAGTAAAAAACAGCTGTGTCCCGTGAATGAGAGCTTCCTTCAATGAGCATGTA 721
OY 758 tgcagacct-cctgagcctggtgagagtgccatctccctcagatgagcagctgagct 816
Db 722 TGCAGACCTCCCTGCAGCTGTGTAAGGTGCATTTTCCCTCAGATGAGCCG-CTGGAGT 780
OY 817 ttgacctctgtatccctgacatcaagagtgacacatcagctctac 863
Db 781 TGGACCTTCTGTGATCTGCATTC-AGGGTGAACCCCATTCAGTTCTAC 826

RESULT 7
LOCUS BG530180 785 bp mRNA linear EST 03-APR-2001
DEFINITION 602358678F1 NIH_MGC.61 Homo sapiens cDNA clone IMAGE:4696842 5',
ACCESSION BG530180 mRNA sequence.
VERSION BG530180.1 GI:13521717
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 785)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNU)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNU at:
http://image.lln.gov
Plate: LCM1524 row: k column: 19
High quality sequence stop: 706.
Location/Qualifiers

```

```

source
1. .785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4696842"
/clone_id="NIH_MGC.61"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: testis; Vector: pBMR-LIB (Clontech); Site:1:
SfiI (ggcgctcgcc); Site:2: SfiI (ggcattatggc);
Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGAGCGGAGCGCCGACATG-dT(30)BN-3'
(where B = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library."
BASE COUNT 170 a 257 c 207 g 151 t
ORIGIN
Query Match 41.8%; Score 702.6; DB 10; Length 785;
Best Local Similarity 97.8%; Pred. No. 2,6e-159;
Matches 764; Conservative 1; Mismatches 10; Indels 6; Gaps 5;
OY 42 gggagagagagggagcgggcccagagatccagcgtgacaggtctgacatcctgactg 101
Db 2 GGGAGAGAGAGAGAGAGGGGCCCGGAGCTCCAGCGTCCAGGCTGGCATCTGCACTTG 61
OY 102 ctgacctcagacacctggaagaatgagccgcgtggagcttcacacctctctgtgttt 161
Db 62 CTGCCCTTGACACCTTGGGAAGATGGCCGGCCGTGGACTTTCACCTTCTGTGTTT 121
OY 162 gctlgagcaccacttgatccaagcaccctcagtcacactgagcttcacatcctgagcc 221
Db 122 GCTGGCAGCACCCTTGATGATCAAGCACCCTCAGTCCCACTGCAAGTTCATCTCGGCC 181
OY 222 aaaaatcacaagaagaagctgacacagagagctgaagagacccaagccacagatcct 281
Db 182 AAAAGTCATCAAGAAAGAGCTGACACAGAGAGCTGAAGAGCACAAAGCCACAGATCCT 241
OY 282 gcaagacgtgcccgtgctcagtgatgcatcgaggaaagcagccgagag-attcctggtct 340
Db 242 GCAGCAGCTGCCGCTGCTCAGTCCATCGGGAAGACCGAGCCGAGCATCTCTGTCT 301
OY 341 gggcagcctggtgtaacacgctcctgaagacatcatctgctgagagtgatcacagctaa 400
Db 302 GGGCAGCCTGTGTAACACCGTCTGAAGCAGTCATCTGGCTGAAGGTCAATCAGCTAA 361
OY 401 catctccagctgacgagtggaagccctcggccaatgacacagagagctgctgaatgacatccc 460
Db 362 CATCTCCAGCTGACAGGTGAAGCCCTCGCAATGACCGAGAGCTGTAGTCAAGATCCC 421
OY 461 cctggaacatggtgctgagatcaacacgcccctgtgtaagacacatggtgagttccat 520
Db 422 CTTGACATGCTGTGCTGATTAACAACGCCCTCGTCAAGCACTCTGGAAGTTCACAT 481
OY 521 gacgaactgagcccaagccacatccgcatggaacacagtgcaatgagccacccgct 580
Db 482 GACGACTGAGGCCAAGCCACCATCCATGGAACACAGGCAAGTGG-CCCAACGCCCT 540
OY 581 ggtcctcagtgactgtgacacagcattggaagcctcgtgacatccaactgctgataagct 640
Db 541 GGTCTCTAGTACTGTGCCACCGCATGGAGAGCTCGCATCAACTGCTGATTAAGCT 600
OY 641 ctctctcctggtgagccttaagcagatgcat--gaacctcagtgacatccctg 698
Db 601 CTCTCTCTGTGTAAGCCCTTACGTAAAGCAGGTCAATGAACCTTCTAGTGCATCCCTG 660
OY 699 ccaatctagtaaaacagct-gtgcctcgatcgaggtcctccttaagga-tgt 756
|||||

```

| Db | 661 | CCGACATCTAGTGAAGAAACAACAGACCTGGTGTGTCGCCGTGATTCAGAGGCTTCCCTTCATGGCATTTGT | 720 |
|------------|-----|--|---|
| QY | 757 | atgcagacctctctgcagcttggtgaaggtgtccatttcctcagcatltagccgtctgagat | 816 |
| | | | |
| Db | 721 | ATGCAGACCTCTCTGCAGCTGGTGAACGGGGGCATTTCCCTCAGACATTCACCGTCTGCAGT | 780 |
| QY | 817 | t | 817 |
| | | + | |
| Db | 781 | T | 781 |
| RESULT | 8 | | |
| BI821769 | | | |
| LOCUS | | 955 bp | mrna |
| DEFINITION | | BI821769 | linear |
| | | 603035838F1 NIH_MGC_115 | EST 04-OCT-2001 |
| | | 603035838F1 | Homo sapiens cDNA clone IMAGE:5176656 5', |
| ACCESSION | | BI821769 | |
| VERSION | | BI821769.1 | |
| KEYWORDS | | GI:15933319 | |
| SOURCE | | EST. | |
| ORGANISM | | human. | |
| | | Homo sapiens | |
| | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| | | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | |
| | | 1 (bases 1 to 955) | |
| | | NIH-MGC | |
| | | http://mgc.nci.nih.gov/. | |
| | | National Institutes of Health, Mammalian Gene Collection (MGC) | |
| | | Unpublished (1999) | |
| | | Contact: Robert Strausberg, Ph.D. | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| COMMENT | | | |

| | | | | | |
|------------|-------|-------|-------|-------|----------|
| BASE COUNT | 216 a | 302 c | 246 g | 190 t | 1 others |
| ORIGIN | | | | | |

| | | | | |
|-----------------------|-----------------|---------------------|-----------|-------------|
| Query Match | 40.8%; | Score 684.6; | DB 10; | Length 955; |
| Best Local Similarity | 93.2%; | Pred. No. 6.4e-155; | | |
| Matches 749; | Conservative 0; | Mismatches 50; | Indels 5; | Gaps 3; |

| Db | 139 | GAAAGCTGACACAGGAGACTGAAAGACCAACAACGCCACACAGATCTTCAGACAGTGGCG | 198 |
|------------|---|---|------------------------|
| OY | 295 | ctgtctcaatgtccatctcgggaaagacagccggag9 -atccctgtgtctgtggcagccctgtg | 353 |
| Db | 199 | CTGCTCAGTGCATCGCGGGAAAGACCGAGGACATCCGTGTCTGGGCGCTGGTG | 258 |
| OY | 354 | aacacgcgtctgaagacacatcatctgtgtgaaggtcatcaagctaacatctctcagctg | 413 |
| Db | 259 | AACACGCTGCTGAACACATCATCTGGCTGGAAGGTCATACAGCTMAATCATCTCCAGCTG | 318 |
| OY | 414 | caagtgaaagccctcgtggccaatgacaggaagctgtcgaftcgaagatctccctctgaaatgtg | 473 |
| Db | 319 | CAGGTGAAGCCCTCGCGCAATGACCAAGAGCGCTGCTAGTCAAGATCCCTCGAGCATGGTG | 378 |
| OY | 474 | gctgattcaacaacgccccctgtgtcaagacacatcgtgtgaggttccacatgaagctgaagcc | 533 |
| Db | 379 | GCTGATTTCAACAGCCGCCCTGTGTCAAGACATCTGTGGAGTTCCACATGACGACTGAGGCC | 438 |
| OY | 534 | caagccacacatctcgcatgagacacagctgcaagtggtgcccaccccgctgttctcagtgac | 593 |
| Db | 439 | CAAGCCACCATTCGCCATGTGAGACACAGTGAAGTGCCACCCGCCCTGTGCTCAGTGAC | 498 |
| OY | 594 | tgtgtccaccagcgcattggagagctgtgcatctcaactgctgcatgaagcttcttctctgtg | 653 |
| Db | 499 | TGTGTCACACAGCCATGGAGCGCTGGGCATTCACAGTGTGACATMAAGCTCTCTCTGGTG | 558 |
| OY | 654 | aacgccttagctaaacagatcagatcagacctctcagtgcatccctgccatctagtga9 | 713 |
| Db | 559 | AACGCTTATAGCTAAACAGAGTCAATGAACCTCTGATGGCATCCCTGCCAATCTACTGAAA | 618 |
| OY | 714 | aacagcgtgtgtcccgtagatcgaagcttctcctcaatgagcatgtatgacgaacctctgca9 | 773 |
| Db | 619 | AACGAGCTGTGTCCCGTATGAGAGGCTTCTTCAATGAGCATGATGACAGACTCTCTGAG | 678 |
| OY | 774 | ctgtggaaggtgtgcca-cttccttagatgtgacgctgtgagtttgacctc---tga | 829 |
| Db | 679 | CTGTGAAGGTGCCCATTTTCCCTAGCATTTGACCGTGTGGAGTTTGACCTTCTTGTATTC | 738 |
| OY | 830 | tctctccatcaaggttgacacccatcaagctctaccctcgtggggccaaagtgtgtgactaca | 889 |
| Db | 739 | CTTGCCATTCMAAGGTGACACCATTTCAAGGCTTACCATGAGGGGGCAGATTGTTCGAGCT | 798 |
| OY | 890 | gggaaagtgaccaaagtgttcaa | 913 |
| Db | 799 | CCACAGCGCAAAAGGCTGACCAA | 822 |
| RESULT | 9 | | |
| LOCUS | B1820886 | 836 bp | mrna |
| DEFINITION | 603033924F1 NIH_MGC_115 Homo sapiens CDNA clone IMAGE:5175388 5', | | linear EST 04-OCT-2001 |
| ACCESSION | B1820886 | | RNA sequence. |
| VERSION | B1820886.1 | GI:15932436 | |
| KEYWORDS | EST. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| TITLE | NIH-MGC http://mgc.nci.nih.gov/. | | |
| JOURNAL | National Institutes of Health, Mammalian Gene Collection (MGC) | | |
| COMMENT | Unpublished (1999) | | |
| | Contact: Robert Strausberg, Ph.D. | | |
| | Email: cgabbs-remail.nih.gov | | |
| | Tissue Procurement: Life Technologies, Inc. | | |
| | cDNA Library Preparation: Life Technologies, Inc. | | |
| | cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) | | |
| | DNA Sequencing by: Incyte Genomics, Inc. | | |
| | Clone distribution: MGC clone distribution information can be | | |
| | found through the I.M.A.G.E. Consortium/LNLN at: | | |
| | http://image.lnl.gov | | |
| | Plate: L1AM11436 | | |
| | row: 0 | column: 05 | |

FEATURES High quality sequence stop: 750.
Location/Qualifiers
source 1..836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:517538"
/clone_1b="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector: PCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."
BASE COUNT 172 a 290 c 210 g 164 t
ORIGIN

Query Match 40.0%; Score 671.6; DB 10; Length 836;
Best Local Similarity 95.2%; Pred. No. 8,4e-152;
Matches 790; Conservative 0; Mismatches 29; Indels 11; Gaps 9;

QY 57 cggcgccgagagctcagcagcgtgccaggtctgcatctgctgctgctgacac 116
Db 1 cggcgccgagagctcagcagcgtgccaggtctgcatctgctgctgacac 60
QY 117 tgggaagatggccggccgtgagacctacacctctctgtgtgtgtgagccacct 176
Db 61 tgggaagatggccggccgtgagacctacacctctctgtgtgtgtgagccacct 120
QY 177 gatcaagccacctgaatccacctgacgttctcactcctcgcgcaaaagatcaaa 236
Db 121 gatcaagccacctgaatccacctgacgttctcactcctcgcgcaaaagatcaaa 180
QY 237 aaagctgaacagagagctggaagagacacaaagccagacatcctgacagcgtcgc 296
Db 181 aaagctgaacagagagctggaagagacacaaagccagacatcctgacagcgtcgc 240
QY 297 gctcagtgccatgctgggaaagccagccgagag-atccctgtgtgtggccgctgtgaa 355
Db 241 gctcagtgccatgctgggaaagccagccgagag-atccctgtgtgtggccgctgtgaa 300
QY 356 cagcgtcctgaagacatcactctgtgtgaagttatcaagcttaacatcctccagctgca 415
Db 301 cagcgtcctgaagacatcactctgtgtgaagttatcaagcttaacatcctccagctgca 360
QY 416 ggtgaa-gccctggccaatgacagagagctgtgag-tcaagatccccctgg--acatgg 471
Db 361 ggtgaa-gccctggccaatgacagagagctgtgag-tcaagatccccctgg--acatgg 420
QY 472 tggctgattcaaacagccctgtgtaaacacatcgtgtgagttccacataagcagcag 511
Db 421 tggctgattcaaacagccctgtgtaaacacatcgtgtgagttccacataagcagcag 480
QY 532 cccaagccacacatccgcatgagacacagttcaagttgcccacccgctggtctcaatg 591
Db 481 cccaagccacacatccgcatgagacacagttcaagttgcccacccgctggtctcaatg 540
QY 592 actg-tgcaacagcagt--ggagcctgcaatccacatgctgctgctgctctcttc 648
Db 541 actg-tgcaacagcagt--ggagcctgcaatccacatgctgctgctgctctcttc 600
QY 649 tggtagagccttaagtagcaggtcattgaacctctagtgccatccctgccaatctag 708
Db 601 tggtagagccttaagtagcaggtcattgaacctctagtgccatccctgccaatctag 660
QY 709 tgaataaacagcgtgtgtccctggtatcgaggtcttcaatgacatgtagtgaagcctcc 768
Db 661 tgaataaacagcgtgtgtccctggtatcgaggtcttcaatgacatgtagtgaagcctcc 720

QY 769 tgcagctgtgaagg-tgccattctccctcagcattgaacgctgtgg-agttgacctct 826
Db 721 tgcagctgtgtgaaggctgcctccatttccctgagccttgcagccttgcagcttgcaccttct 780
QY 827 gtatccttcacatcaagggtgacacacatcagctcactctggtgggccaagt 876
Db 781 gtatccttcacatcaagggtgacacacatcagctcactctggtgggccaagt 829

RESULT 10
LOCUS B548906
DEFINITION 602575739F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4703816 5', mRNA sequence.
ACCESSION B548906
VERSION B548906.1 GI:13547571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 660)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: CLOUTMTECH Laboratories, Inc.
CDNA Library Preparation: CLOUTMTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMU at: http://image.llnl.gov
Plate: LCM1542 row: n column: 09
High quality sequence stop: 660.

FEATURES

source

Location/Qualifiers

1..660
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4703816"
/clone_1b="NIH_MGC_77"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pNIR-LIB (Clontech); Site:1: SfiI (ggcgctcgccg); Site:2: SfiI (ggccatattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCAGAGCGCCGAGCGCGCGCATG-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT 144 a 222 c 176 g 118 t
ORIGIN

Query Match 37.2%; Score 624.2; DB 10; Length 660;
Best Local Similarity 98.9%; Pred. No. 2.2e-140;
Matches 648; Conservative 1; Mismatches 4; Indels 2; Gaps 2;

QY 42 gggagagagatggagggcgccgagagctcagcgtgccaggtctggtcctgcaactg 101
Db 2 gggagagagatggagggcgccgagagctcagcgtgccaggtctggtcctgcaactg 61
QY 102 ctgcccctctgacaccttgggaagatggccggcccggtgagacctcaacctctctgtgtt 161
Db 62 ctgcccctctgacaccttgggaagatggccggcccggtgagacctcaacctctctgtgtt 121
QY 162 gctggagcagcatttatcaagccacctcagttcccaatgacgttctatcctcggccc 221
Db 122 gctggagcagcatttatcaagccacctcagttcccaatgacgttctatcctcggccc 181

| | | | |
|----|-----|--|-----|
| OY | 222 | aaagtcctcaagaagaagctgacacacagagctgtgaagaccacacacgcacagcatcct | 281 |
| Db | 182 | AAAATCTATCAAAAGAAAGCTGACACAGAGCTGGAAGGACCAACAGCCACACAGATCCT | 241 |
| OY | 282 | gcacagactgcgcgtcgtcactgcatgcatgcbgaggaagaacgcagcgagag-atccctgtct | 340 |
| Db | 242 | GCAGCAGCTGCCTGCTGCTCAGTGCCATCGGGAAAAGCAGCGGGAGAGATCCCTGTGCT | 301 |
| OY | 341 | gggcagcctgtgtgaacacgcgttcctgaaagcacatcctgtgctgaaagtcatacagctaa | 400 |
| Db | 302 | GGGCGCCTGGTGAACACCGCTGCTGMAACACACTCATCTGGCTGGAAGGTCATCACAGCTAA | 361 |
| OY | 401 | catctcactgactgcagcagtggaagcccttgagcaatbaacaagagcgtcctagtataaatcc | 460 |
| Db | 362 | CATCTCTCCAGCTGCACGGTGAAGCCTCGGCCATATACAGAGAGCTGTATGTAAATATCC | 421 |
| OY | 461 | cctgagacatggtggtcgtgattcaaacagcccttggtcgaagacatcgttggattccacat | 520 |
| Db | 422 | CCTGGACATGTGTGGCTGTGATTCAACACGCCCTCGTGTCAAGACCATTCGTMGAATTCACAT | 481 |
| OY | 521 | gaagactaaggcccaagcacatcctcgatggaacccaagtgaanttgccccaccgcgt | 580 |
| Db | 482 | GACGACTAAGGCCCAAGCCACCATTCGGATGAGACACCAAGTGAATGTGGCCACCCGGCT | 541 |
| OY | 581 | ggtcctcagtgactgtgcccacacgcacatgaggagcccttgcatcattcaacatgctcataaagt | 640 |
| Db | 542 | GGTCTCTCAGTGACTGTGGCCACACGCCATGGGAGCCTGGGCAT-CAATGCTGCTAATAGGT | 600 |
| OY | 641 | ctcctctcctgtgtgaacgccttagctaaagcaggtcatgaacctcctcctagtgcatacc | 695 |
| Db | 601 | CTCTCTCTCTGTGGAACGCTTATGCTTAACACAGGTCTGTGAACCTCCATGAGTCATCC | 655 |

| FEATURES | source |
|------------|---|
| RESULT 11 | |
| LOCUS | 796 bp mRNA |
| DEFINITION | linear EST 03-APR-2001 |
| ACCESSION | 60256514EF1 NIH_MGC_77 Homo sapiens CDNA clone IMAGE:4693619 5', |
| VERSION | BC540252 |
| KEYWORDS | mRNA sequence. |
| SOURCE | BC540252.1 GI:13532485 |
| ORGANISM | EST. |
| REFERENCE | human. |
| AUTHORS | Homo sapiens |
| TITLE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| JOURNAL | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. |
| COMMENT | 1 (bases 1 to 796) |
| | NIH-MGC http://mgc.nci.nih.gov/ . |
| | National Institutes of Health, Mammalian Gene Collection (MGC) |
| | Unpublished (1999) |
| | Contact: Robert Strausberg, Ph.D. |
| | Email: cgapbs-remail.nih.gov |
| | Tissue Procurement: CLONTECH Laboratories, Inc. |
| | CDNA Library Preparation: CLONTECH Laboratories, Inc. |
| | CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) |
| | DNA Sequencing by: Incyte Genomics, Inc. |
| | Clone distribution: MGC clone distribution information can be |
| | found through the I.M.A.G.E. Consortium/LNLN at: |
| | http://image.lnl.gov |
| | Plate: LNCM1516 row: e column: 12 |
| | High quality sequence stop: 762. |
| | Location/Qualifiers |
| | 1..796 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4693619"
/clone_id="NH_MGC_77"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: lung; Vector: pDNR-LIB (clontech); Site.1:
SfiI (ggcccgcgcgcgc); Site.2: SfiI (ggccatcatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCGCGCGCGCGCGCGCGCAG-dt(30)BA-3' (where B = A,
T, C or G); ATTCTAGAGCGCGCGCGCGCGCGCGCAG-dt(30)BA-3' (where B = A,

```

C, or G and N = A, C, G, or T). Average insert size 1.8 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH/MSC Library.

| | | | | |
|------------|-------|-------|-------|----------|
| BASE COUNT | 178 a | 209 g | 158 t | 1 others |
| ORIGIN | | | | |

| | | | | |
|---------------------------|--------|---------------------|------------|-------------|
| Query Match | 37.1%; | Score 623.2; | DB 10; | Length 796; |
| Best Local Similarity | 94.4%; | Pred. No. 4.2e-140; | | |
| Matches 755; Conservative | 1; | Mismatches 29; | Indels 15; | Gaps 10; |

[illegible]

| RESULT | 12 | | | |
|----------|--------|------|--------|-----------------|
| BG484717 | | | | |
| LOCUS | 912 bp | mRNA | linear | EST 21-MAR-2001 |
| BG484717 | | | | |

| | | | |
|----|-----|--|-----|
| QY | 42 | gggagaggaagtgagagcgagcgaggaatccagagctcccaaggtcttgcaactcgaaactg | 101 |
| Db | 2 | gggagagagagagcgagcgagcgaggaatccagagctcccaaggtcttgcaactcgaaactg | 61 |
| QY | 102 | ctgagccctgaacacctggagaagatgagcgagcgagcgaggaatccagagctcccaaggtcttgcaactcgaaactg | 161 |
| Db | 62 | ctgagccctgaacacctggagaagatgagcgagcgagcgaggaatccagagctcccaaggtcttgcaactcgaaactg | 121 |
| QY | 162 | gctgtagcagcaaccttgatccaaagcaacctcaagtctccactgagtctcatcctctgagcc | 221 |
| Db | 122 | gctgtagcagcaaccttgatccaaagcaacctcaagtctccactgagtctcatcctctgagcc | 181 |
| QY | 222 | aaaagtcatacaagaagaagcttgacacagaagctcgaagagcccaacgcacacgaatcc | 281 |
| Db | 182 | aaaagtcatacaagaagaagcttgacacagaagctcgaagagcccaacgcacacgaatcc | 241 |
| QY | 282 | gcagcgagctgacgctgctcaatgacatgcygggaaaagcagagcgagag-atccctgagct | 340 |
| Db | 242 | gcagcgagctgacgctgctcaatgacatgcygggaaaagcagagcgagag-atccctgagct | 301 |
| QY | 341 | gggacagcctgctgaacaacgctcctgaagacacatcatgctgaaggtcatcaacagctaa | 400 |
| Db | 302 | gggacagcctgctgaacaacgctcctgaagacacatcatgctgaaggtcatcaacagctaa | 361 |

| | | | |
|----|-----|--|-----|
| OY | 401 | caacccacagcttgaagttgaagccctctgcgcacatgatccagagcttgtatgaagatccc | 460 |
| Db | 362 | catctctccagctggaaggggaagccctctggccaatgacacggagctgtatgaatgcc | 421 |
| OY | 461 | ccctgacatgtgtgtctgattcaacacy - gcccttgtcaagacatcgtgtgagttcacca | 519 |
| Db | 422 | ccctgacatgtgtgtctgattcaacagcccccctgtgtcaagacatcgtgtgagttccaca | 481 |
| OY | 520 | tgaagactggaagcccacaagccacatccgcattgtaaccaagt--gcaatgtgtccacaccg | 577 |
| Db | 482 | ttacacactgaaagcccacaaaccacatccgatggatgacacacggttgcaaggttgccacaccg | 541 |
| OY | 578 | ccctgtgctctcagtgactgtgccacagcaatgtgagacccgtgcatacactgtcgtacata | 637 |
| Db | 542 | cctggtctcctcagtgactgtgccacag - catggagacccgtgccatccca - tctgcataag | 599 |
| OY | 638 | gctctcctctctgtgtgaagcctctagct - aagcaagtgatgaactcc--tagtgcat | 693 |
| Db | 600 | ctttctctctctgtgtgaagcctcttacttgaacacagtcataatgtaacccctctgagtgcctatc | 659 |
| OY | 694 | ccctgtcccaactctag - tgaataaacagcctgtgtccg - tttatctagagcttcctctcaatg | 751 |
| Db | 660 | ccctgtcccaactctagttgaataaacacagctgtgtcccggtgattccgagccttcttccatg | 719 |
| OY | 752 | catatagcagactcct - gcaagctgtgtgaaggtgtgccattccctcaagatt - accgt | 809 |
| Db | 720 | catattttgcacagacctctctggcacgtgtgtaaaaggtcccatattttcctcagcaattgaacgt | 779 |
| OY | 810 | ctgagatttaacttc - tttatcctgtccatcaaggttgaaacattc - agctctactgt | 866 |
| Db | 780 | ctgagattttgacatttcctgtattttgtggcattttccaggggtgacacatttcaagttcttacttc | 839 |
| OY | 867 | ggggccaagctgtgtgactccacagggaaagttgaccaagtgattca | 913 |
| Db | 840 | ggggcaattgtttgcattcacaggggacaaaggtcccaagttcttcaaa | 886 |

| | |
|------------|---|
| RESULT | 13 |
| LOCUS | BE061388/c |
| DEFINITION | BE061388 658 bp mRNA linear EST 09-JUN-2000 |
| ACCESSION | QV0-BT0229-251099-038-c01 BT0229 Homo sapiens cDNA, mRNA sequence. |
| VERSION | BE061388 |
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 658) |
| AUTHORS | Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsushima,A., Bais,G.S., Simpson,D.H., Brunstein,A., Dellivelita,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. |
| TITLE | Shotgun sequencing of the human transcriptome with ORF expressed sequence tags |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) |
| MEDLINE | 20020663 |
| COMMENT | Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-27049922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL http://www.ludwig.org.br/scripts/gethtml2.pl?lt=at2-QV0-BT0229-251 099-038-c01&ts=1999-10-25&t4=1) Seq primer: puc 18 forward High quality sequence stop: 620. |

QY 468 atggtgctgagattcaacagccctggtcaagaccatcgtgagttccacatgacgact 527
 DB 423 ATGTTGGCTGATGATTAACAGCCGCCCTGGTCAAGACCATGTGAGTTCACATGACACT 482
 QY 528 ggggcccagccaccatccgcattgacacagtcgaagtggcccccacccgctgtctc 587
 DB 483 GAGGCCCAAGCCACCATCCGATGGACACACTGCAAGTGG-CCCAACCGCTGGTCTC 541
 QY 588 agtgactgtccacagccatggagggcctgcatccactgtctcataagctctcctc 647
 DB 542 ACTGACTGTGCGACCGCATGGAGCCTGGCATTCACACTG-TCGATTAAGCTTCCTTC 600
 QY 648 ctgtgtgaacgcttagctaaagcaggtcattgaacctctctagt 689
 DB 601 CTGTTGAAGCCTTAGCTAAGCAGGTCAAGACTCTCTAGTG 642

RESULT 15

BG482968

LOCUS BG482968 640 bp mRNA linear EST 21-MAR-2001
 DEFINITION 602502981F1 NIH_MGC_77 Homo sapiens CDNA clone IMAGE:4616349 5',
 mRNA sequence.

ACCESSION BG482968
 VERSION BG482968.1 GI:13415247

KEYWORDS EST
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 640)
 NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)

JOURNAL

Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov

COMMENT

Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LUCM1371 row: a column: 22
 High quality sequence stop: 636.

FEATURES

Location/Qualifiers

source

1..640

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4616349"

/clone_lib="NIH_MGC_77"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTATGACC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGGAGCGGCCGCGCATG-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.9
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."

BASE COUNT 140 a 212 c 175 g 113 t
 ORIGIN

Query Match 35.6%; Score 598.8; DB 10; Length 640;
 Best Local Similarity 98.9%; Pred. No. 3, le-134;

Matches 633; Conservative 1; Mismatches 3; Indels 3; Gaps 3;

QY 42 gggagagagagggagcggcgagagactccagctgcccaggtcttgcatctctgacacty 101
 DB 2 GGGAGAGGAGAGGAGCGGCGCGAGACTCCAGAGTCCAGGCTGCCAGGCTGCGATCTGCACTTG 61

QY 102 ctgccccttgacacacctgggaagatggccggccgtggaaccttcaacctctctgtgtt 161
 DB 62 CTGCCCTCTGACACCTGGAGATGGCGGCGCCCTGGAGCTTACCCCTTCTGTGTTT 121
 QY 162 gctggaagcaaccttgatccaaagccacctgaagcccaatgacatcttcaatctctggcc 221
 DB 122 GCTGGACACCACTTGTATCCAAAGCCACCGCTGATGCCATGTCATGCTTCCTGGGCC 181
 QY 222 aaaaatcaaaaagaaagctgacacagagctgaagagacacaacagcaacacacacac 281
 DB 182 AAAAGTATCAAAAGAAAGCTGACACAGAGAGCTGAAGAGCACACAGCCACACAGCACT 241
 QY 282 gcaacacatgcccgtctgtctgaatgctgaggggaaagcaagccaggagg-atccctgtgtc 340
 DB 242 GCAGCAGCTGCGGCTGCTCAAGTGCATGCGGAAAGCCAGCGAGGCAATCCCTGTGCT 301
 QY 341 gggcagcctgtgagacacccgtcctgaagacacatcatctgtgaagttcaacacacaa 400
 DB 302 GGGCAGCCTGTGTGAACACCGTCTGAAAGCAGTATCTGGCTTAAGTCAATCAAGCTAA 361
 QY 401 catctccagctgcaaggtgaagccctcgccaatgacaaagagctgtatgaagatccc 460
 DB 362 CATCTTCACCTGAGGTGAAGCCCTCGGCCAATGACACAGAGCTGTACTAAGATCC 421
 QY 461 cctggacatgtgtgctgattcaacagccccctggtcaagacatctgtgagttcaacat 520
 DB 422 CTGAGACATGTGTGCTGTGATTCAACAGCCCTGTGTAAAGCATGTGAGTTCACAT 481
 QY 521 gacgactgaagcccaagccaccaaccgcatgagacacagtgcaagtggccccacccgct 580
 DB 482 GAGGACTGAGGCCCAAGCCACCATCCGATGAGACACAGTGCAGAGTGG-CCAGCGCGCT 540
 QY 581 ggtctcagtgactgtgcccacagccatggagctgagcatcgaactgtgataaact 640
 DB 541 GGTCTCAGTGTGTGCCACGACCATGGAGCCCTGCCATCCAACTGCTGATTAAGCT 600
 QY 641 ctcttctgtgtgaac-gccttagctaaagagttatga 679
 DB 601 CTCCTTCCTGTGATGAGCGGCTTATGCTAAGCAGGTCTATGCA 640

Search completed: September 16, 2002, 01:42:43
 Job time: 3296 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 02:57:19 : Search time 3785.3 Seconds
(without alignments)
4947.891 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895

Sequence: 1 ctaacgttactgtaacagc.....aataaatttttaagtgc 895

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_fun:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_un:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | ID | Description |
|------------|-------------|--------|----|-------------|
|------------|-------------|--------|----|-------------|

| | | | | | | |
|----|-------|-------|--------|----|----------|-------------|
| 1 | 895 | 100.0 | 895 | 6 | AX224644 | Sequence |
| 2 | 895 | 100.0 | 1171 | 6 | AX224648 | Sequence |
| 3 | 852 | 95.2 | 870 | 6 | AX092294 | Sequence |
| 4 | 525.8 | 58.7 | 533 | 6 | AX062486 | Sequence |
| 5 | 475 | 53.1 | 626 | 6 | AX182041 | Sequence |
| 6 | 438.4 | 49.0 | 168586 | 6 | AC011497 | Homo sapi |
| 7 | 414.8 | 46.3 | 419 | 6 | AX332462 | Sequence |
| 8 | 357 | 39.9 | 357 | 6 | E40581 | Novel prote |
| 9 | 312 | 34.9 | 312 | 6 | AX062416 | Sequence |
| 10 | 291 | 32.5 | 291 | 6 | E40584 | Novel prote |
| 11 | 245.6 | 27.4 | 357 | 6 | E40583 | Novel prote |
| 12 | 228 | 25.5 | 357 | 6 | E40582 | Novel prote |
| 13 | 200.4 | 22.4 | 291 | 6 | E40586 | Novel prote |
| 14 | 190 | 21.2 | 44332 | 6 | AC005794 | Homo sapi |
| 15 | 189.2 | 21.1 | 281 | 6 | E40585 | Novel prote |
| 16 | 138.2 | 15.4 | 40520 | 9 | AC011513 | Homo sapi |
| 17 | 118.4 | 13.2 | 43593 | 9 | AC022516 | Homo sapi |
| 18 | 78.8 | 8.8 | 190860 | 2 | AC073803 | Mus muscu |
| 19 | 78.8 | 8.8 | 215234 | 2 | AC073745 | Mus muscu |
| 20 | 73.4 | 8.2 | 97 | 6 | AX182077 | Sequence |
| 21 | 56.8 | 6.3 | 7218 | 6 | I66494 | Sequence 14 |
| 22 | 45.8 | 5.1 | 518 | 6 | AX185399 | Sequence |
| 23 | 44.6 | 5.0 | 51 | 6 | E40591 | Novel prote |
| 24 | 44 | 4.9 | 125020 | 9 | AF429315 | Homo sapi |
| 25 | 44 | 4.9 | 209887 | 2 | AF429315 | Mus muscu |
| 26 | 43.6 | 4.8 | 125020 | 9 | AF429315 | Mus muscu |
| 27 | 43.2 | 4.8 | 234787 | 2 | AC073780 | Mus muscu |
| 28 | 42.8 | 4.8 | 456 | 6 | AX187025 | Sequence |
| 29 | 42.4 | 4.7 | 209703 | 2 | AL645853 | Mus muscu |
| 30 | 41.4 | 4.6 | 163392 | 2 | AC097208 | Rattus no |
| 31 | 41.2 | 4.6 | 67356 | 9 | AL161658 | Human DNA |
| 32 | 40.6 | 4.5 | 170212 | 9 | AC058822 | Homo sapi |
| 33 | 40.6 | 4.5 | 198614 | 10 | AL596127 | Mouse DNA |
| 34 | 40.6 | 4.5 | 211844 | 2 | AC048370 | Homo sapi |
| 35 | 39.4 | 4.4 | 209156 | 2 | AL607088 | Mus muscu |
| 36 | 39.2 | 4.4 | 205150 | 9 | AC005274 | Homo sapi |
| 37 | 39 | 4.4 | 167159 | 2 | AC073481 | Homo sapi |
| 38 | 39 | 4.4 | 196416 | 9 | AC006483 | Homo sapi |
| 39 | 38.6 | 4.3 | 7218 | 6 | I66494 | Sequence 14 |
| 40 | 38.6 | 4.3 | 152719 | 2 | AC012345 | Homo sapi |
| 41 | 38.6 | 4.3 | 155949 | 9 | AC091022 | Homo sapi |
| 42 | 38.6 | 4.3 | 156863 | 2 | AC069249 | Homo sapi |
| 43 | 38.6 | 4.3 | 161775 | 9 | AC099512 | Homo sapi |
| 44 | 38.6 | 4.3 | 187574 | 9 | AC018622 | Homo sapi |
| 45 | 38.6 | 4.3 | 213049 | 2 | AC079370 | Mus muscu |

ALIGNMENTS

RESULT 1
AX224644 LOCUS AX224644 895 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 2 from Patent WO0161055.
ACCESSION AX224644
VERSION AX224644.1 GI:15554773

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Chen, S.Y., Sun, Y. and Macina, R.A.
Methods for diagnosing, monitoring, staging, imaging and treating
lung cancer via lung cancer specific genes
Patent: WO 0161055-A 2 23-AUG-2001;
diadexus, Inc. (US)

FEATURES

source

1. 895
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 245 a 250 c 183 g 217 t
ORIGIN

Db 742 CTCCTCTAGTGTCTTCTCTGTCAGTCCTAGCTTGCCCTCCCTTACCCAGGCT 801
Qy 781 taggcttaattacctgaaagatltccagaaactgtagcttcttagtgcatttaac 840
Db 802 TAGGCTTAATTAACCTGAAAGATTCCAGGAACGTGAGCTTCTAGCTAGTGCATTAAAC 861
Qy 841 cttaaatgcaatcagaaagtagcaaaacagaagtcataatatltttaagctc 895
Db 862 CTTAATGCAATCAGAAAGTAGCAACAGCAAGTCATAATATTTTAAATGTC 916

RESULT 3
AX092294
LOCUS AX092294 870 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 25 from Patent WO0116318.
ACCESSION AX092294
VERSION AX092294.1 GI:13444461
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Eaton,D.L., Filvaroff,E., Gerlitsen,M.E., Goddard,A.,
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
Wood,W.I.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 25 08-MAR-2001;
Genentech, Inc. (US)
FEATURES
source 1..870
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 251 a 237 c 176 g 206 t
ORIGIN

Query Match 95.2%; Score 852; DB 6; Length 870;
Best Local Similarity 100.0%; Pred. No. 2.1e-254;
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 ctgcgcctcaaaatggaacgtgcgtggaactaaagcagacacacagagctgagtc 103
Db 1 CTCGGCCCTCAATATGGGAACGCTGGCCTGGACTAAAGCATACCAACGAGCTGAGTATC 60
Qy 104 ctgacctgagtcaccccaaggatcaagagcctccaagcaggaacaccttcattatct 163
Db 61 CTGACCTGAGTCATCCCCAGGGATCAGAGCCTCCAGCAGGGAACCTTCATTATATCT 120
Qy 164 tcaagaacttcaagctgacagcagatgagtgatgaagttctaattcttccctctcc 223
Db 121 TCAAGCAACTTAAGCTGACGACGACAGTGGATGAATTAATCTTCCCTCTCTCC 180
Qy 224 tgttgctgcacataatgctgcatgcatgctcttagcagcctgaaatccaagggtcgc 283
Db 181 TGTTCCTGCACATTAATGCTGATGCTCATGCTCTCTAGCAACCTGAAATCCAGGGCTGCCA 240
Qy 284 gaggcacacaggaacccgagcaggtcttagagatgctcacaaggaaggcgccaagaat 343
Db 241 GAGGCCACAGGGAGCGAGCGCAGGCTTCTAGAGATGCTCCAGGAAGCGCGCAAGAT 300
Qy 344 gtgaagtcaaaagattggttctctgagagcccgagaaagaatattcatgacagtgtctgg 403
Db 301 GTGAGTGCAGAAATGTGGTCTGAGAGCCCGAGAGAAATTCAGAGACGTCTCTGGGC 360
Qy 404 tgcacaaagaagcagtgccctgtgcatcttcaagggcaatgtgaagaaacaagacacc 463
Db 361 TGCAGAAAGAGCAGTCCCTGTGATCATTTCAAGGGCATGTGAAGAAACAGACAC 420
Qy 464 aagaagcacacagaaagcaaaacaaagcaltccagagcctgcagcaatlttcaaaat 523
Db 421 AAAGGACACCAAGAAAGCCAAACAGCATTTCCAGAGCCTGCCAGCAATTTCTCAACAAAT 480

Qy 524 gtcacaaagaagacttctgctgccttctgtagagctcagcagccacttccaat 583
Db 481 GTACGCTAAGAGCTTTGCTGTGCTTGTAGAGCTCTGAGCGCCACTCTTCCAATTA 540
Qy 584 aacattctcagcacaagacagtagcacacctaacagacactcttctccacactc 643
Db 541 AACATTCTCAGCAGAAAGACAGTAGCGACACTACACACTCTCTCTCCACCTC 600
Qy 644 actctccacatgtaacccacccctaataatccagtgctctcaaaaagcattttcaa 703
Db 601 ACTCTCCACACTGTACCAACCCCTAAATCAATTCAGTGTCTCAAAAAGCATTTTTTCAA 660
Qy 704 gatacttgttattgttcctctcctcagtgctctctctctctctctcagcttcaacgtgac 763
Db 661 GATCATTTTGTGTGTGCTCTCTAGTGTCTTCTCTCTGTCAGTCTTACCTGTGCTC 720
Qy 764 ctcccttaccacagccttagcttaattacctgaaagattccaaggaactgtgcttct 823
Db 721 CTCCTCTACCGCAGGCTTAGGCTTAATTAATTCGAAAGATTCCAGGAACCTGTAGCTTCT 780
Qy 824 agctagtgcatltaaccttaaatgcaatcaggaagtagcaaacagaaagtcataata 883
Db 781 ACCTAGTGTCAATTTAACCCTTAATGCATCAGAAAGTAGCAACAGAGTCATAATAATA 840
Qy 884 ttttaaatgctc 895
Db 841 TTTTAAATGTC 852

RESULT 4
AX062486
LOCUS AX062486 533 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 113 from Patent WO0100828.
ACCESSION AX062486
VERSION AX062486.1 GI:12540361
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
Carter,D., Reltter,M.W. and Mannion,J.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0100828-A 113 04-JAN-2001;
CORIXA CORPORATION (US)
FEATURES
source 1..533
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 152 a 154 c 113 g 114 t
ORIGIN

Query Match 58.7%; Score 525.8; DB 6; Length 533;
Best Local Similarity 99.6%; Pred. No. 1.2e-152;
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 179 ctgcacgacagttgagatgaagtctaattcttccctccctcctgctgagccactaa 238
Db 1 CTGCACCGACAGTGTGGATGAAGATTCTTAATCTTCCCTCCTCTGTTGCGCCACTAA 60
Qy 239 tgcatalgtcaltgtctccttagcagcctgaatccaggggtgcagagagcccaaggacc 298
Db 61 TCTGTATGTCCATGTGCTCTACAGCCTGAATCCAGAGGCTCCGACAGAGCGAC 120
Qy 299 gaggcagagcctctaggaatgctcctcagaagaggcgccaagaatgtgagtgcaaat 358
Db 121 GAGGCCAGGCTTCTAGAGATGCTCCAGAAAGCGCGCCCAAGATTTGAGTGAAGAAATTT 180
Qy 359 ggttccttgagagcccgagagaagaatltcatgacagtgtctgggtcgcacaaagacagt 418

| | | | |
|----|-----|---|-----|
| XY | 57 | ggagcgtggtccctgggactaaagcatagacacagctgatactcctgacctgaftca | 116 |
| XY | 35 | gggaagcttgccctggggacttAAAGCTAAAGCACCAAGGGTGGAGTCTCGACTGAGTCA | 94 |
| XY | 117 | tccccaggatcagaagccctccagcagggaaaccttcacatattatcttcaagcaattac | 176 |
| XY | 95 | TCGCCAGGATCAGAGACCCTCCAGCAGGGAACTTCATTTATATCTTCAAGCAACTTAC | 154 |
| XY | 177 | agctgacccgaacagttgcgatgaagttcttaattcttcctcctcctcgttgcgcact | 236 |
| XY | 155 | AGCTGCACCGACAGTGGCGATGAATAATTCTATCTCTCTCCCTCCGTTGTGTCGCACT | 214 |
| XY | 237 | aatgtcatgtcatgctctctagcagcttgataatcaagggttcgacagaagcccaagga | 296 |
| XY | 215 | AATGCTGATGTCATGGTCTCTAAGACGCTGAAATTCAGAGGGTGGCAAGGCCACAGGGA | 274 |
| XY | 297 | ccgaagcccaagctctcaagatagatgctccagaagcgcgccaagaatgtgagtcaaga | 356 |
| XY | 275 | CCGAGGCCAGGCTTTCAGAGAGATGGCTCCAGCAAGCGCCCAAAATGTGTGATGTCGAAGA | 334 |

[illegible]

Db 70597 TCCAAATTAAACATTCCTCAGCCCAAGACAGTGCACACCTACACGACACTCTTCTCT 70656
Qy 636 cccacctcaactcctccactgctacccacccttaactcattccagtgctctccaaaagatg 695
Db 70657 CCACCTCTCATCTCCACAGTACCCACCCCAATTCATTCAGTCTCTCAAAAAGCAAG 70716
Qy 696 ttcttcaagaatcatttgctgtctctctctagtgctctctctctctctctctctctctag 755
Db 70717 TTTTTCAGATCATTTTGTGTTTCTCTCTCTAGTGTCTTCTCTCTCTCTCTCTCTG 70776
Qy 756 cctgtgcctccctccctaccagagcttaggcttaattaccctgaagaattccaggaactgt 815
Db 70777 CCTGTGCCCTCCCTTACCAGGCTTAGGCTTAATTACCTGAAGAATTCAGGAACCTGT 70836
Qy 816 agcttcctagctagtgatcatttaacttaaatgcaatcaggaagaatagaaacaagaatgc 875
Db 70837 ACCTTCTCAGCTAGTGTCTCTTAACTTAAATGCAATGCAAGAAATGCAAGCAAGTGC 70896
Qy 876 aataaatattttaaatgctc 895
Db 70897 AATAAATATTTTAAATGTC 70916

RESULT 7
LOCUS AX332462 419 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 2971 from Patent WO0194629.
ACCESSION AX332462
VERSION AX332462.1 GI:18123096
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrikan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 2971 13-DEC-2001;

FEATURES
source Avalon Pharmaceuticals (US)
Location/Qualifiers
1..419
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 121 a 63 c 117 g 118 t
ORIGIN

Query Match 46.3%; Score 414.8; DB 6; Length 419;

Best Local Similarity 99.5%; Pred. No. 5.2e-118;

Matches 416; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 478 aagcaaacaaagcattccagagcctgacgaatcttccaacaaatgltcagtaagaagc 537
Db 419 AAGCCAACAAAGCATTCAGAGCTGCGCAGCAATTTCTCAACAAATGTCAGTAAGAAGC 360
Qy 538 ttgtctctgctcttgtagagctctgagcgcccaactcttccaattaacaattcagaaca 597
Db 359 TTTGCTCTGCTTTGTAGAGCTGTGAGGCGCCACATTCATTAACATTCCTCAGCCA 300
Qy 598 agaagaacagtagcaacactaccagaacactcttctcccaacctcaactcccaactgta 657
Db 299 ACAAAGACAGTAGACACACTACCAAGACACTCTTCTCTCCACCTCAGCTCCACACTGTA 240
Qy 658 cccaccccttaaatcattccagtgctctcaaaaagcatgttttccaagaatcatttgctg 717
Db 239 CCACCCCTTAATTCATTCAGTGCCTCTCAAAAAGCATGTTTTCAGATCATTTCTTGTG 180
Qy 718 ttgtctctctctagtgctctctctctctctcagtccttagccttgccctccctaccag 777
Db 179 TTGCTCTCTCTAGTGTCTTCTCTCTGCTGACGTTCAGCTTGCCCTCCCATTAACCCAG 120

Qy 778 gcttagcttaattaccctgaagaattccaggaactgtagcttctctagctagtgatct 837
Db 119 GCTTAGGCTTAATTACCTGAAGAATTCAGGAAGTGTAGCTTCTAGTATGTTCATTT 60
Qy 838 aaccttaaatgcaatcaggaagtagcaacagaaagtcataataattttaaatgctc 895
Db 59 AACCTTAATGCAATCAGGAAGTGTGCAAGCAAGAGTCAATTAATTTTAAATGTC 2

RESULT 8
LOCUS E40581 357 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel protein and process for producing it.
ACCESSION E40581
VERSION E40581.1 GI:18625110
KEYWORDS JP 2000139479-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 357)
AUTHORS Ito,T., Ogi,K., Tanaka,H. and Kitada,C.
TITLE Novel protein and process for producing it
JOURNAL Patent: JP 2000139479-A 1 23-MAY-2000;
TAKEDA CHEM IND LTD
OS Homo sapiens (human)
PN JP 2000139479-A/1
PD 23-MAY-2000
PF 02-SEP-1999 JP 1999248414
PR

COMMENT
PI YASUAKI ITO,KAZUHIRO OGI,HIDEYUKI TANAKA,CHIEKO KITADA PC
C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00, PC
A61K31/00,
PC A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
PC A61K38/00,
PC A61K45/00,A61K48/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,
PC G01N33/50//
PC A61K39/395,C12P21/08,(C12P21/02,C12R1.91),C12N15/00,A61K37/02
CC
FH key
FT source
Location/Qualifiers
1..357
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 102 a 92 c 90 g 73 t
ORIGIN

Query Match 39.9%; Score 357; DB 6; Length 357;

Best Local Similarity 100.0%; Pred. No. 5.6e-100;

Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 196 atgaaagttctaattctctctcctcctctctctctctctctctctctctctctctctctct 255
Db 1 ATGAAAGTTCTAATCTCTTCT 60
Qy 256 tctagcagcttgaaatccagaggtgcgcagaggtccaaaggaagcagaggtccagctctag 315
Db 61 TCTAGACGCTGAATCCAGAGGTCGCGCAGAGGCGACAGAGGCGAGGCGCAGGCTTCTAG 120
Qy 316 agatgctcctcaggaagggcgcaagaatgtagtgcaagaatgtagtgctctgagagcccg 375
Db 121 AGATGCTCTCAGAGGCGCGCGCAAGATGTGATGCAAGATTGCTTCTCTGAGAGCCCG 180
Qy 376 agaagaanaattatcacagtgtctgaggtgcgaagaagcaggtgcccctgtatcatctc 435
Db 181 AGAAGAAATTTATGACATGTCTGTGGCTGCCAAGAAACAGTACGCTGTGATCTATTTC 240
Qy 436 aagggcaatgtgaagaaacaaagacaccaaaggtcaccacagaagaagcacaagaatcc 495
Db 241 AAGGGCAATGTGAAGAAACAAAGACACCAAGGCGACCAAGAAAGCAAGCAATTC 300

QY 496 agagcctgcagcaattctcaacaatgcaagagcttgccttg 552
|||||
Db 301 AGAGCCTGCAGCAATTCTCAACAATGCTAGCTAAGAGCTTGGCTGCTTTG 357

RESULT 9
AX062416 312 bp DNA linear PAT 24-JAN-2001

LOCUS AX062416
DEFINITION Sequence 43 from Patent WO0100828.
ACCESSION AX062416
VERSION AX062416.1 GI:12540291
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 312)
AUTHORS Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
Carter,D., Retter,M.W. and Mannion,J.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0100828-A 43 04-JAN-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..312
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 100 a 82 c 69 g 61 t
ORIGIN

Query Match 34.9%; Score 312; DB 6; Length 312;
Best Local Similarity 100.0%; Pred. No. 6e-86;
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 caggagcgagccagaatgtgagtgcaaatgtgttcctgagagcccgagagaagaaa 384
|||||
Db 1 CAGGAGGCGGCGCAAGATGTGACTGCAAGATTGTTCTTGAGAGCCCGAGAGAAAA 60

QY 385 ttctgacagtgctgtggtgcgaagaagcaatgtccctgtgatacttcaagggcaat 444
|||||
Db 61 TTCTATGACAGTGTCTGGCTGCGCCAAAGAGCAGTCCCTGTGATCTATTCAAGGGCAAT 120

QY 445 gtgagaagaacaagacacccaagcaccacagaagaagcaagaacattccaagctgc 504
|||||
Db 121 GTGAGAGAAAACAACACCAAGCCACACAGAGCAAAACAGATTCCAGAGCCTGC 180

QY 505 cagcaattctcaacaatgtcagctaagaagcttgcctgcttgaagagctctga 564
|||||
Db 181 CAGCAATTCTCAACAACATGTCAAGTGAAGAGCTTGGCTTGTGAGAGCTCTGA 240

QY 565 ggcgccactcttcgaatgaacattctcagccagaagaagacagtgagcaacactaccagac 624
|||||
Db 241 GCGGCCACTTCTTCCAAATTAACATTCTCAGCCAAAGAGACAGTGAACCTACCCAGAC 300

QY 625 actcttctctc 636
|||||
Db 301 ACTCTTCTTCTC 312

RESULT 10
E40584 291 bp DNA linear PAT 31-JAN-2002

LOCUS E40584
DEFINITION Novel protein and process for producing it.
ACCESSION E40584
VERSION E40584.1 GI:18625113
KEYWORDS JP 2000139479-A/4.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 291)

AUTHORS Ito,Y., Ogi,K., Tanaka,H. and Kitada,C.
TITLE Novel protein and process for producing it
JOURNAL Patent: JP 2000139479-A 4 23-MAY-2000;
TAKEDA CHEM IND LTD
COMMENT OS Homo sapiens (human)
PN JP 2000139479-A/4
PD 23-MAY-2000
PF 02-SEP-1999 JP 1999248414
PR

PI YASUAKI ITO,KAZUHIRO OGI,HIDEYUKI TANAKA,CHIERO KITADA PC
C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,PC
A61K31/00,
PC A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
PC A61K38/00,
PC A61K45/00,A61K48/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,
PC G01N33/50//
CC A61K39/395,C12P21/08,(C12P21/02,C12RL:91),C12N15/00,A61K37/02
CC
FH
FT
FT Key source Location/Qualifiers
1..291
/organism="Homo sapiens (human)".
1..291
Location/Qualifiers
1..291
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 90 a 72 c 79 g 50 t
ORIGIN

Query Match 32.5%; Score 291; DB 6; Length 291;
Best Local Similarity 100.0%; Pred. No. 2.1e-79;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 agcttaattccaggggtgcgcagagggccagggagccagagcttctagagatg 321
|||||
Db 1 AGCTTAATTCACAGGGGTGCCAGAGGCCACAGGGACCGAGCGCAGCTCTAGAGATGG 60

QY 322 ctccaagagcgcgccaagaatgtgagtgcaaatgtgttcttgaagagcccgagaaga 381
|||||
Db 61 CTCAGAGAGGGCGGCCAATAATGTGAGTGCATAAGATTGGTTCTTAGAGCCCGAGAGAA 120

QY 382 aaattcaagcagtgctgtggtgcgaagaagcagtgccctgtgatacttcaagggc 441
|||||
Db 121 AATTCTATACAGTGTCTGGCTGCCAAAGAGCAGTGCCTGTGATCATTTCAAGGGC 180

QY 442 aetgtgaagaacaagacacccaagcaccacagaagaagcaagaacattccaagacc 501
|||||
Db 181 AATGTGAAGAAAACAAGACACCAAGGCCACACAGAAAGCCAAACAGCATTTCCAGAGCC 240

QY 502 tgcagcaattctcaacaatgtcagctaagaagagcttgcctgcttg 552
|||||
Db 241 TCCAGCAATTTCTCAACAATGTGAGCTAGAGAACTTGGCTGCTGCTTTG 291

RESULT 11
E40583 357 bp DNA linear PAT 31-JAN-2002

LOCUS E40583
DEFINITION Novel protein and process for producing it.
ACCESSION E40583
VERSION E40583.1 GI:18625112
KEYWORDS JP 2000139479-A/3.
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 357)
AUTHORS Ito,Y., Ogi,K., Tanaka,H. and Kitada,C.
TITLE Novel protein and process for producing it
JOURNAL Patent: JP 2000139479-A 3 23-MAY-2000;
TAKEDA CHEM IND LTD
OS Mus sp. (mouse)
PN JP 2000139479-A/3
PD 23-MAY-2000


```

PC      A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
PC      A61K38/00,
PC      A61K45/00,A61K48/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,
PC      G01N33/50//
PC      A61K39/395,C12P21/08,(C12P21/02,C12N1/91),C12N15/00,A61K37/02
CC
FH      Key
FH      source
FT      Location/Qualifiers
       1..291
       /organism="Mus sp. (mouse)"
BASE COUNT      88 a      82 c      80 g      41 t
ORIGIN
Query Match      22.4%; Score 200.4; DB 6; Length 291;
Best Local Similarity 80.7%; Pred. No. 4e-51;
Matches 234; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY      262 agcctgaatccaggggtgcagagagccacagagccagagccttctagagatgg 321
Db      1 ACCCCGAACCCAGGGGGTCCCGAAGCCAGCGGACCAACACTGCTCTAGAGGTGG 60
QY      322 ctccaggaagcgccgaagaatgtgaaatgtgttctcctagagagcccgagaaga 381
Db      61 CTTCTGGAAGGTGGCCAGAAATGTGAATGCAAAAGATGTGCTCTCAAGCCCAAGAGA 120
QY      382 aatcagaacagtgctcgtgcagagagcagtcctctgtatcatcttcaaggc 441
Db      121 AAAGCCACAGCAGTCTGGGGCCACCAAGAAAGCGTCTCCCTGATGACGTCAAGGGC 180
QY      442 aatgtgaagaagaacagacaccagaagcaccagaagaagccaagaacattcagaagc 501
Db      181 AGGGGAAAAAACAAGACACCAAAAGCACCAGACAGACGCAAAAGCCCTCCAGAGCC 240
QY      502 tgcagcaatttccaacaacatgtcagtaagaagcttgcctgcctt 551
Db      241 TGCCAGCAATTTCACAAACGATGTACCTGGCAGCTTTGGCGCCCTT 290

RESULT 14      44332 bp      DNA      linear      PRI 06-Oct-1998
AC005794      Homo sapiens chromosome 19, cosmid F24866, complete sequence.
LOCUS      AC005794
DEFINITION      AC005794
ACCESSION      AC005794.1 GI:3702272
VERSION      HTG.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 44332)
AUTHORS      Lamerdin,J.E., McCready,P.M., Skowronski,E., Vistwanathan,V.,
      Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stillwagen,S.,
      Phan,H., Velasco,N., Do,L., Regala,W., Ferry,A., Gerns,U.,
      Dangnan,L., Exler,A., Christensen,M., Georgescu,A., Avila,J.,
      Liu,S., Atlix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
      Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
      Kromayashl,B., Arellano,A., Saunders,C., Ow,D., Nolan,M., Trong,S.,
      Sequence analysis of a 2.5 Mb region in 19q13.2 containing a
      clustered CEA/PSG gene family
      Unpublished
      2 (bases 1 to 44332)
      Lamerdin,J.E.
      Direct Submission
      Submitted (06-Oct-1998) Joint Genome Institute, Lawrence Livermore
      National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
      Map and sequence oriented from q centromere to telomere. Cosmid
      F24866 overlaps cosmid R32065 to the left from bases 1 to 5,578 of
      this accession, and overlaps F13825 to the right from bases 35,645

```

```

FEATURES
      source
      to 44,332. Additional chromosome 19 map and sequence information
      are available from: http://www-bio.llnl.gov/genome/genome.html.
      Location/Qualifiers
      1..44332
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="19"
      /map="19q13.2 between BCKDHA and D19S217"
      /clone="F24866"
      /cell_line="GV5HR9-5B"
      /clone_id="HL19NC02 F chromosome 19-specific cosmid
      library"
      /note="Cosmid library constructed at LLNL from flow-sorted
      chromosomes from hybrid 5H2-B, which carries chromosome
      19 as its only human chromosome."
      599..674
      /rpt_family="MIR"
      complement(1521..1691)
      /note="predicted exon, program: grail2exons_human_1.3,
      frame: 1, quality: excellent, score: 84.000"
      complement(join<1522..1880,2814..2974))
      /product="Human CGM7 gene for nonspecific cross-reacting
      antigen (NCA)"
      complement(<1522..2877)
      /gene="CGM7"
      /note="CGM7"
      /codon_start=1
      /product="nonspecific cross-reacting antigen W236
      precursor - human [AA 1-141]"
      /protein_id="AAC62829.1"
      /db_xref="GI:3702273"
      /translation="MGPPSAAPRGGRHPGGLITLASLTFEHPPTVOFTLEALPSS
      AAEQDVLLACNISEITQAYYHKGKTAEGSPILAGYITDQANIPEAAISGNETYV
      PMSGLFQNTLEDAGSYLRITNNSYSDQATGQLHGH"
      complement(1984..2039)
      /rpt_family="(CA)n"
      2063..2382
      /rpt_family="LINE2"
      complement(4877..6389)
      /rpt_family="MER65 Internal"
      complement(6739..7153)
      /rpt_family="MER65A"
      7570..7673
      /rpt_family="LMC/D"
      complement(7805..8107)
      /rpt_family="AluDb"
      complement(8288..8583)
      /rpt_family="AluSg"
      complement(8955..9420)
      /rpt_family="MER51B"
      complement(9721..9927)
      /rpt_family="L1PA7"
      complement(9919..12991)
      /rpt_family="L1"
      13161..13329
      /rpt_family="MSTA"
      13331..13604
      /rpt_family="AluSx"
      13613..13831
      /rpt_family="MSTA"
      complement(11573..15821)
      /note="BLASTX similarity to (151..233); match: 0.57,
      score: 4.2e-23; database searched: nr; biliary
      glycoprotein h precursor - human g11179435 (M69176)
      biliary glycoprotein a [Homo sapiens] >gn1PDI01015048
      (D90312)"
      complement(15573..15821)
      /note="BLASTX similarity to (151..233); match: 0.57,
      score: 4.7e-23; database searched: nr; biliary
      glycoprotein 1 precursor - human g11179437 (M72238)
      misc_feature

```

```
misc_feature
complement(15573..15800)
/note="BLASTX similarity to 37204 (248..322); match: 0.32,
score: 1.0e-06; database searched: nr; (X16356) TW5-CEA
protein [Homo sapiens]"
complement(15612..15722)
/note="predicted exon, program: gra112exons_human_1.3,
frame: 2, quality: good, score: 58.000"
repeat_region
/rpt_family="MER49"
16142..16326
/rpt_family="AluIo"
complement(16418..16513)
/note="BLASTX similarity to 437784 (3..34); match: 0.65,
score: 7.2e-08; database searched: nr; (221818)
carcinoembryonic antigen [Homo sapiens]"
repeat_region
/rpt_family="FRAM/FAM"
17280..17362
/rpt_family="MIR"
17560..17662
/rpt_family="MIR"
17666..17801
/rpt_family="AluIo"
17864..17958
/rpt_family="MIR"
17973..19083
/rpt_family="L1MC4"
19174..19461
/rpt_family="AluY"
complement(19462..19504)
/rpt_family="AT_rich"
19543..19624
/rpt_family="MIR"
19786..20086
/rpt_family="AluX"
complement(20324..20730)
/rpt_family="L1PB2"
complement(20855..21143)
/rpt_family="AluDb"
complement(21173..21261)
/rpt_family="MIR"
21341..21542
/rpt_family="MER20"
complement(22570..22669)
/note="BLASTN similarity to M87707 (105..204); match: 1,
score: 6.2e-33; database searched: nr; Human simple repeat
polymorphism."
complement(22675..22723)
/rpt_family="(CA)n"
22976..23187
/note="predicted exon, program: gra112exons_human_1.3,
frame: 1, quality: good, score: 62.000"
repeat_region
complement(23278..23613)
/rpt_family="LTR8"
complement(23863..24002)
/rpt_family="LTR8"
complement(24872..25091)
/rpt_family="MER4_internal"
complement(25088..25368)
/rpt_family="MER57_internal"
complement(25389..25768)
/rpt_family="MER57_internal"
26147..26317
/note="BLASTX similarity to Q13946 (7..63); match: 0.78,
score: 5.0e-30; database searched: nr; HIGH-AFFINITY
CAMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE (HCPI1)"
26320..26427
/note="BLASTX similarity to Q13946 (65..100); match: 0.63,
score: 5.0e-30; database searched: nr; HIGH-AFFINITY
CAMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE (HCPI1)"
26474..26668
/rpt_family="L1M1/2"
26666..28684
repeat_region
```

```
/rpt_family="L1M1/2"
28652..30135
/rpt_family="L1"
repeat_region
complement(30136..30213)
/rpt_family="THE1A"
30219..36641
/rpt_family="L1PA7"
complement(36659..36969)
repeat_region
/rpt_family="THE1A"
complement(36970..37380)
/rpt_family="THE1A-internal"
complement(37381..37656)
/rpt_family="THE1B"
37657..39326
/rpt_family="L1PB1"
complement(39711..40104)
/rpt_family="LTR8"
complement(40124..40360)
/rpt_family="L1PA5"
complement(40626..40743)
/rpt_family="LTR8"
40791..40952
/rpt_family="MIR"
complement(41469..42015)
/note="BLASTN similarity to A0108164 (2..548); match:

Query Match      21.2%; Score 190; DB 9; Length 44332;
Best Local Similarity 84.9%; Pred. No. 1.6e-47;
Matches 236; Conservative 0; Mismatches 40; Indels 2; Gaps 2;

QY 1 ctaactgtctacgtacacaa-gacagcgtacaccctcctcgtctcgcctcaatc 59
DB 23226 CTAATCTGTTACTTAACGATTAAGCCAGCATACCTCACCTGCTCTGCCCTCAGATGGA 23167

QY 60 aacgctggcctggagactaaagcatagaccacagcggtagatcctgacctagtcac 119
DB 23166 AACACTGACCTGGAGACTAAAGTGACAGACCAGCGGCTGAGTACCCGTGAGCCATCC 23107

QY 120 caagagatcagaagcctccagcaggaaccttcattatattcttcaagaacttaacgc 179
DB 23106 CCAGAGATCAGGAGCCCTCAACCGGGAACCTTCA-TACAGTCTCAATGACTTACACG 23048

QY 180 tgcaccagagcttgcgataaagttctatctctccctccctcgttgcgcacatc 239
DB 23047 TGCACCAACAATCACATTAAGTTCTATCTCTTCCCTCCTCGTGTCTGCCACTAAT 22988

QY 240 gctgatlctcagtlctctagcagcctgaatccaggg 277
DB 22987 GCTAATGACCATGTCCTTAGCAGCCCAAGTCCAGGTG 22950

RESULT 15
LOCUS E40585 291 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel protein and process for producing it.
ACCESSION E40585
VERSION E40585.1 GI:18625114
KEYWORDS JP 2000I39479-A/5.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 291)
AUTHORS Ito,Y., Ogi,K., Tanaka,H. and Kitada,C.
TITLE Novel protein and process for producing it
JOURNAL Patent: JP 2000I39479-A 5 23-MAY-2000;
TAKEDA CHEM IND LTD
COMMENT OS Rattus sp. (rat)
PN JP 2000I39479-A/5
PD 23-MAY-2000
PR 02-SEP-1999 JP 1999248414
```

PI YASUAKI ITO, KAZUHIRO OGI, HIDEYUKI TANAKA, CHIEKO KITADA PC
 C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00, PC
 A61K31/00,
 PC A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00,
 PC A61K38/00,
 PC A61K45/00, A61K48/00, C07K14/47, C07K16/18, C12P21/02, G01N33/15,
 PC G01N33/50//
 A61K39/395, C12P21/08, (C12P21/02, C12P21/91), C12N15/00, A61K37/02
 CC
 FH
 FT
 Key Location/Qualifiers
 source 1. .291
 Location/Qualifiers
 1. .291
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 BASE COUNT 90 a 79 c 78 g 44 t
 ORIGIN

Query Match 21.1%; Score 189.2; DB 6; Length 291;
 Best Local Similarity 78.3%; Pred. No. 1.3e-47;
 Matches 227; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
 QY 262 agcctgaatcgaaggggtcgccagagggccacagggagccgaagccaggtctctaagagatgg 321
 Db 1 AGCCCGAATCAAGAGGTGCGCACACACCATGGGATCAACACACAGCTCCTAGAGAGGTGG 60
 QY 322 ctccaaggaagggcgccaagaatgtgaatgcaagaattgttctctgagagcccgagaaga 381
 Db 61 CTCTGGGAGGTGCGCAAGAGTGTGACTGCAGAGATTGTTCTCTGCGAGTCTCAAGAGA 120
 QY 382 aaatcattgacagtgtctggtgctgccaagaagcagtgccctgtgtatcatltaaaagc 441
 Db 121 AAAACACACAGCAGTGTGAGGACCAAGAGAGCAGTGTCTGTGATCATGTCAAGGC 180
 QY 442 aatgtgaagaagaac 501
 Db 181 AGTAGAAGAAAGAACAGACGCAAAAGACACAGAGTCAACAAAGGCCCTCCAGAAC 240
 QY 502 tgcacgaattctcaacaattgcaagaattgcaagaagcttctctgcctt 551
 Db 241 TGCACGAATTTCTCAAGCATGTCAACTAGCAAGCTTCGCCCTGCCTT 290

Search completed: September 16, 2002, 02:58:21
 Job time: 7714 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:31:07 : Search time 3813.96 Seconds
(without alignments)
4910.710 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895
Sequence: 1 ctaatcgttaccgaacagc.....ataaatattttaatgtc 895

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | DB ID | Description |
|------------|-------------|--------|-------|-------------|
|------------|-------------|--------|-------|-------------|

| | | | | | | |
|----|-----|-------|--------|----|-----------|-------------|
| 1 | 895 | 100.0 | 895 | 6 | AX224644 | Sequence |
| 2 | 895 | 100.0 | 1171 | 6 | AX224648 | Sequence |
| 3 | 852 | 95.2 | 870 | 6 | AX092294 | Sequence |
| 4 | 474 | 53.0 | 626 | 6 | AX182041 | Sequence |
| 5 | 438 | 48.9 | 168586 | 9 | AC011497 | Homo sapi |
| 6 | 427 | 47.7 | 533 | 6 | AX062486 | Sequence |
| 7 | 357 | 39.9 | 357 | 6 | E40581 | Novel prote |
| 8 | 316 | 35.3 | 419 | 6 | AX332462 | Sequence |
| 9 | 312 | 34.9 | 312 | 6 | AX062416 | Sequence |
| 10 | 291 | 32.5 | 291 | 6 | E40584 | Novel prote |
| 11 | 68 | 7.6 | 97 | 6 | AX182077 | Sequence |
| 12 | 42 | 4.7 | 456 | 6 | AX187025 | Sequence |
| 13 | 42 | 4.7 | 518 | 6 | AX185599 | Sequence |
| 14 | 40 | 4.5 | 51 | 6 | E40591 | Novel prote |
| 15 | 37 | 4.1 | 40520 | 9 | AC011513 | Homo sapi |
| 16 | 36 | 4.0 | 43593 | 9 | AC022516 | Homo sapi |
| 17 | 31 | 3.5 | 47 | 6 | E40590 | Novel prote |
| 18 | 28 | 3.1 | 291 | 6 | E40586 | Novel prote |
| 19 | 28 | 3.1 | 357 | 6 | E40583 | Novel prote |
| 20 | 28 | 3.1 | 190860 | 2 | AC073803 | Mus muscu |
| 21 | 28 | 3.1 | 215234 | 2 | AC073745 | Mus muscu |
| 22 | 26 | 2.9 | 40 | 6 | E40588 | Novel prote |
| 23 | 26 | 2.9 | 40 | 6 | E40589 | Novel prote |
| 24 | 22 | 2.5 | 74433 | 2 | AC015098 | Drosophi1 |
| 25 | 22 | 2.5 | 75470 | 2 | AC036159 | Homo sapi |
| 26 | 22 | 2.5 | 152786 | 2 | AP000878 | Homo sapi |
| 27 | 22 | 2.5 | 168514 | 3 | AC022350 | Drosophi1 |
| 28 | 22 | 2.5 | 177577 | 3 | AC022349 | Drosophi1 |
| 29 | 21 | 2.3 | 300 | 14 | AF409528 | HIV-1 Iso |
| 30 | 21 | 2.3 | 13364 | 4 | AF171666 | Bos tauru |
| 31 | 21 | 2.3 | 62871 | 9 | AL356860 | Human DNA |
| 32 | 21 | 2.3 | 67340 | 2 | AC100905 | Mus muscu |
| 33 | 21 | 2.3 | 101961 | 9 | AL133335 | Human DNA |
| 34 | 21 | 2.3 | 113431 | 2 | AC094336 | Rattus no |
| 35 | 21 | 2.3 | 125413 | 2 | AC026607 | Homo sapi |
| 36 | 21 | 2.3 | 128277 | 2 | AC069107 | Homo sapi |
| 37 | 21 | 2.3 | 137127 | 2 | AC069194 | Homo sapi |
| 38 | 21 | 2.3 | 149623 | 2 | AP004693 | Oryza sat |
| 39 | 21 | 2.3 | 156042 | 2 | AC013614 | Homo sapi |
| 40 | 21 | 2.3 | 164396 | 2 | AC024961 | Homo sapi |
| 41 | 21 | 2.3 | 164396 | 2 | AC024961 | Homo sapi |
| 42 | 21 | 2.3 | 170751 | 2 | AC015742 | Homo sapi |
| 43 | 21 | 2.3 | 180868 | 2 | AL672040 | Homo sapi |
| 44 | 21 | 2.3 | 183624 | 30 | AC019101 | Homo sapi |
| 45 | 21 | 2.3 | 189389 | 9 | CNS050TCE | Human chr |

ALIGNMENTS

RESULT 1
AX224644 LOCUS AX224644 895 bp DNA 11linear PAT 10-SEP-2001
DEFINITION Sequence 2 from Patent WO0161055.
ACCESSION AX224644
VERSION AX224644.1 GI:15554773

SOURCE

ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 895)

REFERENCE

AUTHORS Chen,S.Y., Sun,Y. and Macina,R.A.
TITLE Methods for diagnosing, monitoring, staging, imaging and treating lung cancer via lung cancer specific genes
JOURNAL Patent: WO 0161055-A 2 23-AUG-2001;
diabexus, Inc. (US)

FEATURES

Source

1. 895
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 245 a 250 c 183 g 217 t
ORIGIN

| | | | | |
|-----------------------|-----------------|---------------|-----------|-------------|
| Query Match | 100.0%; | Score 895; | DB 6; | length 895; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 895; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|--|-----|
| QY | 1 | ctaatcgtltaagtaacagcaagacagcgctcaactcaactcgtlcttcgcctcaaatggga | 60 |
| Db | 1 | CTAATCTGTTACCTTAACAGCAAGACAGCGCTCACTCACTGTTCTTCGGCCTCAATGGGA | 60 |
| QY | 61 | acgcctggccctggagcaataagcattagaccacacagcgctgagtaacctgactgagtaacc | 120 |
| Db | 61 | ACGCTGGCCTGGAGACTTAAGACATAGACACACAGCGCTGATATCTCGACTGTAGTATCC | 120 |
| QY | 121 | caaggatcaggaagcctccagccagggagacctcattatatacttcctaagaacattacgct | 180 |
| Db | 121 | CAGGATCAGGAGCGCTCCAGCAGGAGCACTTCATATATTTCTTCAGCACTTACACT | 180 |
| QY | 181 | gaacgaagcttggagatgaaatgttcaatcttccctccctcctcctgtgtgcgactaaag | 240 |
| Db | 181 | GAACGACAGCTTGCAATGAAATTTCTAATCTTCCCTCCCTCTGTTGCTCGCACTAATG | 240 |
| QY | 241 | ctgatgtccatgtcctctagcagccttgaaatccaggggtgcgcagagggccagagagccga | 300 |
| Db | 241 | CTGATGTCCATGSGTCTCTAGCAGCGCTGAATCCAGGGGTGCCAAGGCCACAGGACCGA | 300 |
| QY | 301 | ggccagagcctctaggagatggtctccaaggaagcggcgaagaatgtgagtgcaaaatttg | 360 |
| Db | 301 | GGCCAGGCTCTTAGAGATGGCTCCAGGAAGCGGCCAAGATGTGATGTGCAAAAGATTGG | 360 |
| QY | 361 | ttccctgagagcccgagagaagaatctcatgaaagtcttggtgcctgcgaagaagcaatgc | 420 |
| Db | 361 | TTCCCTGAGAGCGCCGAGAAAGAAATTCATGACAGTGTGGGGCTGCCAAAGACAGTGC | 420 |
| QY | 421 | cccttgaatcatttaagaaggcaatgttgaagaagaacaagacaccaaagcaccacagaag | 480 |
| Db | 421 | CCCTTGATCATTTTCAAGGGCAATGTGAAGAAACAAACACCAAAGGCACACAGAAAG | 480 |
| QY | 481 | ccaaaacaagcattccagaagcctgcgcgaagaattctccaagaagtgtcagctaaagaagttt | 540 |
| Db | 481 | CCAAACAAGCATTTCCAGAGCGTGGCAGCAATTTCCAAACAATGTCACTTAGAGGCTTT | 540 |
| QY | 541 | gctcggccttggtagagcctctgagcgcccaactctccaattaaacaattctcagccaga | 600 |
| Db | 541 | GCTCGCCTTGTGAGAGACTCTGAGCGGCCCACTTCCAAATTAACAATTCACACCAGA | 600 |
| QY | 601 | agaacagtgaacaacctacccagaacactcttctcccaactgaactctccoaagtaacc | 660 |
| Db | 601 | AGACAGTAGACACACTTACAGACAGACTCTTTCTCCCACTCACTCTCCACAGTACCC | 660 |
| QY | 661 | accctcaaatcaatccagtgctctcaaaaagcatgttttccaagatcaatttggtttg | 720 |
| Db | 661 | ACCCCTAATATCTTCAGTGTCTCCAAAAGCATGTTTTCAGATCATTTTGTGTGG | 720 |
| QY | 721 | ctctctcagtgctctctctctctgtaagcttagagctgtagccctcccttaccagagct | 780 |
| Db | 721 | CTCTCTCTAGTGTCTTCTTCTCTGTAGCTTTTAGCTGTGGCTCCCTTACCAGGCT | 780 |
| QY | 781 | tagccttaattaccttgaaagaattccagaaactgttagctcttcctcagctagctgatttaac | 840 |
| Db | 781 | TAGCCTTAATTACTCGAAAGATTCCAGGAACCTAGCTCTTAGGTAGTGCTATTAAAC | 840 |
| QY | 841 | cttaatgcaatcaggaagtagaacaacgaagtcataatattttaatgtc | 895 |
| Db | 841 | CTTAATGCAATCAGGAAGTAGAACAACAGAGTCATAAATTTTAAATGTCT | 895 |

| RESULT | 2 | | | | |
|------------|------------|--------------------------|---------|-----|-----------------|
| AX224648 | | | | | |
| LOCUS | AX224648 | | 1171 bp | DNA | |
| DEFINITION | Sequence | 6 from Patent WO0161055. | | | |
| ACCESSION | AX224648 | | | | |
| VERSION | AX224648.1 | GI:15554775 | | | PAT 10-SEP-2001 |

| | |
|------------|---|
| KEYWORDS | human. |
| SOURCE | |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. |
| AUTHORS | 1 (bases 1 to 1171) |
| TITLE | Chen, S.Y., Sun, Y. and Macina, R.A. |
| JOURNAL | Methods for diagnosing, monitoring, staging, imaging and treating lung cancer via lung cancer specific genes |
| FEATURES | Patent: WO 0161055-A 6 23-AUG-2001; |
| source | diadexus, Inc. (US) |
| | Location/Qualifiers |
| | 1..1171 |
| | /organism="Homo sapiens" |
| | /db-xref="taxon:9606" |
| BASE COUNT | 333 a 320 c 225 g 293 t |
| ORIGIN | |

| | | | | |
|-----------------------|-----------------|---------------|-----------|--------------|
| Query Match | 100.0%; | Score 895; | DB 6; | Length 1171; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 895; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |

[illegible]

| Db | 742 | CTCTCTCTAGTGTCTTTCTTCTCTGTGTAGCTCTTACGCTGTGTGCGCTCCCTCTTACCCAGGCT | 801 |
|-----------------------|---|---|-----------------|
| Oy | 781 | tagccttaattaccgcgaagaattccaggaactctgagcttcctagctagtgtaattac | 840 |
| Db | 802 | TAGGCTTATTATACCTGGAAGATTCCAGGAACCTGTAGCTTCTTACGTGTGTCATTTAAC | 861 |
| Oy | 841 | cttaaatgcattcaggaagttagcaacaagaagtcataataatttttaaatgtc | 895 |
| Db | 862 | CTTAATATCAATTCAGGAAGTAGCAACAGAAAGTCAATTAATATTTTTAAATGTC | 916 |
| RESULT | 3 | | |
| AX092294 | AX092294 | 870 bp | DNA |
| LOCUS | Sequence 25 from Patent WO0116318. | | linear |
| DEFINITION | AX092294 | | PAT 21-MAR-2001 |
| ACCESSION | AX092294 | | |
| VERSION | AX092294.1 | | GI:13444461 |
| KEYWORDS | | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 870) Eaton,D.,U., Flivaroff,E., Gerritsen,M.E., Goddard,A., Godowski,P.,J., Grimaldi,C.,J., Gurney,A.L., Watanabe,C.K. and Wood,W.I. | | |
| TITLE | Secreted and transmembrane polypeptides and nucleic acids encoding the same | | |
| JOURNAL | Patent: WO 0116318-A 25 08-MAR-2001; Genentech, Inc. (US) | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..870 /organism="Homo sapiens" /db_xref="taxon:9606" | | |
| BASE COUNT | 251 a 237 c 176 g 206 t | | |
| ORIGIN | | | |
| Query Match | 95.2%; Score 852; DB 6; Length 870; | | |
| Best Local Similarity | 100.0%; Pred. NO. 0; | | |
| Matches | 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | |
| Oy | 44 | ctcgccctcaaatggaagcgtcgctgagtaagcatagaccacagcgtgagtc | 103 |
| Db | 1 | CTCGGCCCTCAATATGGGAACGCTGGCTGGAGCTAAGCATAGACCAACAGCGTGAGTATC | 60 |
| Oy | 104 | ctgacctgagtcattccaccaggaatcagaagcctccagcaggaagccttcattatct | 163 |
| Db | 61 | CTGACCTCAAGTCATCCACGAGGATCAGGAGCGCTCCAGCAGGGGAACCTTCATTATATCT | 120 |
| Oy | 164 | tcaagaacttaccagctgcacagcaagtttgagtaaaagtttaattcttcctccctcc | 223 |
| Db | 121 | TCAACCAACTTACACCTGCACTGCACGACAGTGTGATAAAGTTCTATCTTCTCCCTCC | 180 |
| Oy | 224 | tgttcctgaccatactgtatgltccatcagtgcttcacagccttgatccaggygtcgca | 283 |
| Db | 181 | TGTTGCTGCCACTATATGCTGATGTCATAGTGTCTTAGCAGCCTGAAATCAGGGGTGCGCA | 240 |
| Oy | 284 | gagccacacaggaaccgaagcaggtctctagaagatggtccagaagcagcgcaagaat | 343 |
| Db | 241 | GAGGCCACAGGAGCCGAGGCGCGCTTCTAGAGAGATGCTCCAGGAAAGCGCGCAAGAT | 300 |
| Oy | 344 | gtgaagtgaagaattggtctcctgaaagcccgagagaagaattcatgacagtgctgggc | 403 |
| Db | 301 | GTAAGTGCAGAAAGATTGTCTCTGAGAGGCCCGAGAGAAATTCATGACAGTGTGGGC | 360 |
| Oy | 404 | tgcacaagaagcagtgccctgtgatacttcaagaggaatgtgaaagaacaagacaccc | 463 |
| Db | 361 | TGCCAAAGAGAGTGGCCCTGTGATCAATTTCAAGGCAATGTGAAGAAACAAACACACC | 420 |
| Oy | 464 | aaagcaccacagaagaaccacaagaattccagaagcctgcagcaatttccaagaat | 523 |
| Db | 421 | AAAGCACCCACGAAGGCCAAACAGCAATTCAGAGCGCTGCCAGCAATTTTCAACACAT | 480 |

| | | | |
|---|------------|---|-----------------|
| OY | 524 | gtcagctaagaagcttctgtcttctgtttagtgagctctgagccacactcttccaatla | 583 |
| Db | 481 | gttcacctaaagaagcttctgtctgtcctttgttagagactctgagccacactcttccaatla | 540 |
| OY | 584 | aacattctcagcaagaagaagtagagacacactcacaagaactcttctcccactc | 643 |
| Db | 541 | AACATTCTCAGCCAAGAAGACAGTAGACACACTTACACAGACTTTCTTCTCCACCTC | 600 |
| OY | 644 | actctcccaactgtaccacacccctaaatcaltccacagctgtctctcaaaaagatgttttcaa | 703 |
| Db | 601 | ACTCTCCCACTGTATCCACACCCCTAATATCATTCGATGCTCTCAAAAAGCATGTTTTCAA | 660 |
| OY | 704 | gatacttltgttctgtctctctctagtgctctcttctctctgtcagctttagctgtgcc | 763 |
| Db | 661 | GATCATTTTGTGTGTGTGCTCTCTCTAGTGCTCTTCTTCTCTGCTGAGTCTTGA6CCTGTG6CC | 720 |
| OY | 764 | ctccccttaccagagcttagagcttaacttaaccctggaagaattccagaagaactgtagcttct | 823 |
| Db | 721 | CTCCCTTACCCAGGAGCTTAGGCTTAATTAATCTGGAAGAATTCAGAGAAACTGTAGCTTCTCT | 780 |
| OY | 824 | agcttagtgcatttaccattaaatgtcaatcaggaagaatgacaaagaagtaataa | 883 |
| Db | 781 | AGCTAGTGTGATTTAACTTAATATCATTCAGGAAGAATGAGCAAAAGATCAATTAATA | 840 |
| OY | 884 | ttttaaatgtc | 895 |
| Db | 841 | TTTTTAAATGTC | 852 |
| RESULT 4 | | | |
| AX182041 | LOCUS | AX182041 | 626 bp |
| AX182041 | DEFINITION | Sequence | DNA |
| AX182041 | ACCESSION | AX182041 | linear |
| AX182041.1 | VERSION | AX182041.1 | PAT 06-AUG-2001 |
| AX182041.1 | KEYWORDS | GI:15133314 | |
| SOURCE | ORGANISM | human. | |
| | | Homo sapiens | |
| | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| | | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| REFERENCE | AUTHORS | 1 (bases 1 to 626) | |
| | TITLE | Lillie,J., Palermo,A., Wang,Y., Steinmann,K. and Elias,J. | |
| | | Identification, assessment, prevention, and therapy of breast | |
| JOURNAL | | Cancer | |
| | | Patent: WO 0146697-A 51 28-JUN-2001; | |
| FEATURES | SOURCE | Millennium Predictive Medicine, Inc. (US) | |
| | | Location/Qualifiers | |
| | | 1..626 | |
| | | /organism="Homo sapiens" | |
| | | /db_xref="taxon:9606" | |
| BASE COUNT | | 172 a 170 c 149 g 133 t | 2 others |
| ORIGIN | | | |
| Query Match 53.0%: Score 474; DB 6; Length 626; | | | |
| Best Local Similarity 100.0%: Pred. NO. 1.6e-261; Indels 0; Gaps 0; | | | |
| Matches 474; Conservative 0; Mismatches 0; | | | |
| OY | 57 | gggaacgcctgagcctgagactaaagataagacaccacagagctgagtaacctgactgagta | 116 |
| Db | 35 | GGGAACGCGTGGCTGGGACTAAAGCATATGAGACACAGAGGTGATTCCTGACACTGATCA | 94 |
| OY | 117 | tcaccaagagatcaggagcctccagcagaggaaccttcacatatacttccaagaactaac | 176 |
| Db | 95 | TCACCAGGAGATCCAGAGCCTCCAGCAGGGAACCTTCCATATATTTCTTCAMCAACTTAC | 154 |
| OY | 177 | aagctcaccagaaagctgcagataagaagtcttaactcttcccccctccgtgtccact | 236 |
| Db | 155 | AGCTCAGCGACAGATTGGCATGAAGTTCTTAATCTCTTCCCTCCTCTGTTGCTGCACT | 214 |
| OY | 237 | aatgtctatgttccatggtctcttagacagctgtaatccaggagtcgcacaagagccacagga | 296 |

Db 215 AATGCTGATGTCATGCTCTCTAGCAGCCTGAATCCAGGGGTGCGCAGAGGCCACAGGGA 274
 QY 297 ccgagccagagctcttagagatggtccaggaagcgcccaagaatgtgagtgaaga 356
 |||||||
 Db 275 CGAGGCCAGGCTTTAGAGATGGCTCCAGGAGGCGCCAGAAATGTGAGTGCAGAA 334
 QY 357 ttggttcctgaagaccccgagaagaatctatgacagtgtctggctgccaagaaga 416
 |||||||
 Db 335 TTGGTTCCTGAGAGCCCGAGAGAAATTCATGACAGTGTCTGGGCTCCAAAGAGCA 394
 QY 417 gtgcctctgatactttaaagggaatgttgaagaacaagacaccaaagccacacag 476
 |||||||
 Db 395 GTGCCCTGTGATCATTTTCAAGGGCAATGTGAAGAAACACCAAGCAGCACCACAG 454
 QY 477 aaagccaaacagcatctcagagcgctgcagcaatcttccaacatgtcagct 530
 |||||||
 Db 455 AAAGCCAAACAGCATTTCCAGAGCCTGCCAGCAATTTCTCAACATGTCACT 508

RESULT 5

LOCUS AC011497 168586 bp DNA linear PRI 30-JUN-2000
 DEFINITION Homo sapiens chromosome 19 clone CTB-50E14, complete sequence.
 AC011497
 VERSION AC011497.6 GI:8844110
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 168586)
 DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE

Direct Submission

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 168586)
 DOE Joint Genome Institute.

AUTHORS

Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

JOURNAL

Submitted (30-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT

On Jun 30, 2000 this sequence version replaced gi:7711512.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov

FEATURES

Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 100% of Sequence;
 Estimated Total Number of Errors is 0.

source

Location/Qualifiers
 1..168586
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /clone="CMB-50E14"

BASE COUNT 44482 a 43600 c 40985 g 39515 t
 ORIGIN

Query Match 48.9%; Score 438; DB 9; Length 168586;
 Best Local Similarity 100.0%; Pred. No. 7.7e-241;
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 458 gacaccaaagacacacagaagaacacaaagcatccagagcctgcagcaattctca 517
 |||||||
 Db 70479 GACACCAAAAGCAGCAGACAGAAACCAACAGCATCCAGAGCCTCCAGCAATTTCTCA 70538
 QY 518 aacaatgtaagcaagaagctgtcctcctctgttagagagctcgaagcgccacatctc 577
 |||||||
 Db 70539 AACAAATGTCAGCTTAAGAGGCTTGTGCTCTTTGAGAGGCTGTGAGCGCCACACTCTTC 70598

QY 578 caatlaaacatctcagccaagaagacagtgagcacacactaccagacactctctcc 637
 |||||||
 Db 70599 CAATTAACATTTCTCAGCCAGAGAGACAGTGAGCACACACTTCAAGACACTTTCTTCC 70658
 QY 638 caccacacatccacacatgtaccacaccccaataatcaccagtgctctcaaaagatgt 697
 |||||||
 Db 70659 CACCTCAGCTCCACAGCTGTACCCACCCCTTAATATTCACATGCTCTCAAAACCATGTT 70718
 QY 698 ttcaagaatcatttggctgtgtctctcctcagctgtctctctcctcagcttagcc 757
 |||||||
 Db 70719 TTTCAGATCATTTTGTGTTGTGCTCTCTAGTGTCTTCTCTCTGTGTAAGTTAGCC 70778
 QY 758 ttgtccctcccttaaccagagcttagcttaactacatgaagaattccaggaactgtag 817
 |||||||
 Db 70779 TGTGCCCTCCCTTACCAGCAGCTTAGGCTTAATTACCTGAAAGATTCAGAGAACTGTAG 70838
 QY 818 ctctcagtagtgatcatttaaccttaatgcaatcaggaagaagagaacagaagca 877
 |||||||
 Db 70839 CTTCTAGCTAGTGTCTATTTAACCTTAATGCAATCAGGAAAGTAGCAACAGAGTCAA 70898
 QY 878 taatatatttaaatgctc 895
 |||||||
 Db 70899 TAAATATTTTAAATGTC 70916

RESULT 6

LOCUS AX062486 533 bp DNA linear PAT 24-JAN-2001
 DEFINITION Sequence 113 from Patent WO0100828.
 AX062486
 ACCESSION AX062486
 VERSION AX062486.1 GI:12540361
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 533)
 Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedtick,T.S.,
 Carter,D., Retter,M.W. and Mannin,J.
 Compositions and methods for the therapy and diagnosis of lung
 cancer

AUTHORS

Patent: WO 0100828-A 113 04-JAN-2001;
 CORIXA CORPORATION (US)

JOURNAL

Location/Qualifiers
 1..533
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

FEATURES

source

BASE COUNT 152 a 154 c 113 g 114 t
 ORIGIN

Query Match 47.7%; Score 427; DB 6; Length 533;
 Best Local Similarity 99.6%; Pred. No. 2.1e-234;
 Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 179 ctgcacccagagctgcagatgaagatctatctctccctcctctgttgcgcacaa 238
 |||||||
 Db 1 CTGCACCGACAGTTGGAGATGGCTCCAGAAAGCGGCCAAGATGTGCAAGATT 180
 QY 239 tgcgatgtcattgtgtctctcagagcctgaatcagaggggtgcagaaagccagagacc 298
 |||||||
 Db 61 TCGTAGTGCATGTGCTCTTACACCTGAAATCCAGGGGTGCGCAGAGGCCACAGGAGCC 120
 QY 299 gaggcagagctctcagagatagctcccgagaagggcgccaagaatgtgagtcaagaatt 358
 |||||||
 Db 121 GAGGCCAGGCTTTTAGAGATGGCTCCAGAAAGCGGCCAAGATGTGCAAGATT 180
 QY 339 gttcctcgaagcccgagaagaatcagagagtgctcgggctgcgaagaagcagtt 418
 |||||||
 Db 181 GATTCCTGAGAGCCCGAGAGAAATTCATGACAGTGTCTGTGGCTGCCAAAGAGCACT 240
 QY 419 gccctgtgacatcttcaagggcaatgtgaagaacaagacaccaaagggcacacagaa 478
 |||||||

Db 241 GCCCTGTGATCATTTCAAGGGCAATGTGAGAAAAACAAGACCAACCAAGCACACAGAA 300
 QY 479 agccaaacagatcaccagagcctccagcaatttcacaacagtcagtaagact 538
 Db 301 ACCCAACAGAGATCCCAAGAGCTGCACCAATTTCTCAACAAATGTCAGTAAAGACT 360
 QY 539 ttgctctgcttctgttagagagctcgaagcgcacactcttcacattaaacttcagccaa 598
 Db 361 TTGCTCTGCTTGTGTAGAGCTGTGAGGCCCACTCTTCAATTAACAATTTCTCAAGCCAA 420
 QY 599 gaagacagtgagacacacacacacacactctctctccacactcactccacactgac 658
 Db 421 GAAGACAGTGAGACACACACACACACACTCTTCTCCACACTCAGCTGACACTGATAC 480
 QY 659 ccaccccaataatcattccagtgctctcaaaaagcatgttttcagatc 707
 Db 481 CCACCCCTAAATCATTCACAGTGTCTCAAAAAGCATGTTTTCAGATGC 529

RESULT 7
 EA0581
 LOCUS E40581 357 bp DNA linear PAT 31-JAN-2002
 DEFINITION Novel protein and process for producing it.
 ACCESSION E40581
 VERSION E40581.1 GI:18625110
 KEYWORDS JP 2000139479-A/1.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ito.Y., Ogi.K., Tanaka.H. and Kitada.C.
 Novel protein and process for producing it
 Patent: JP 2000139479-A 1 23-MAY-2000;
 TAKEDA CHEM IND LTD
 OS Homo sapiens (human)
 PN JP 2000139479-A/1
 PD 23-MAY-2000
 PF 02-SEP-1999 JP 1999248414

REFERENCE
 AUTHORS Ito.Y., Ogi.K., Tanaka.H. and Kitada.C.
 TITLE Novel protein and process for producing it
 JOURNAL Patent: JP 2000139479-A 1 23-MAY-2000;
 TAKEDA CHEM IND LTD
 COMMENT
 OS Homo sapiens (human)
 PN JP 2000139479-A/1
 PD 23-MAY-2000
 PF 02-SEP-1999 JP 1999248414

PI YASUAKI ITO, KAZUHIRO OGI, HIDEYUKI TANAKA, CHIEKO KITADA PC
 C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00, PC
 A61K31/00,
 PC A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00,
 PC A61K38/00, A61K48/00, C07K14/47, C07K16/18, C12P21/02, G01N33/15,
 PC A61K45/00, A61K48/00, C07K14/47, C07K16/18, C12P21/02, G01N33/15,
 PC G01N33/50//
 PC A61K39/395, C12P21/08, C12P21/02, C12R1:91), C12N15/00, A61K37/02
 CC
 FH
 FT
 KEY source location/Qualifiers
 FT source location/Qualifiers
 FEATURES
 source location/Qualifiers
 1..357
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 102 a 92 c 90 g 73 t
 ORIGIN

Query Match 39.9%; Score 357; DB 6; Length 357;
 Best Local Similarity 100.0%; Pred. No. 5,2e-194;
 Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 196 atgaagttcttaactctctcctcctcctctgtctgtgcacataatgctgaatgcatgctc 255
 Db 1 ATGAAGTCTTAATCT 60
 QY 256 tcttagagctgaatcagagagtgctgagcagagcagagcagagcagagcagagcagagcagagc 315
 Db 61 TCTAGAGGCTGTAATCTCAAGGGGCGCCAGAGGCCACAGGAGGCCAGGCGCATGCTTTAGG 120
 QY 316 agatgctcagagagcgccagagatgtgagtgcaagatgtgtcctgagagcccg 375

Db 121 AGATGGCTCAGAGAGCGCGCCAGAAATGTAGTCTCAAAAGATTTGGTCTCTGAGAGCCCG 180
 QY 376 agaagaaatcattagcagtgctgtgctgtgcacaaagcagtgccctgtgacatc 435
 Db 181 AGAAGAAATTCATGACATGCTCTGGGGTGCCTAAAGAACATGCTGCCCTGTATCATTTG 240
 QY 436 aaggagcatgtgaaagaaacaaagacacaaagcaccacacagaaagcacaagcattcc 495
 Db 241 AAGGGCAATGTAGAGAAAAACAAGACCAAAAGGACACACAGAAACCAAAAGCATTTGC 300
 QY 496 agagctgtcagcaattctcacaacatgtcagctaaagagctgtgctgtcctg 552
 Db 301 AAGGCTGCCACATTTCTCAACAAATGTCACTAAGAGCTTGTCTGCTGCTTTG 357

RESULT 8
 AX332462/c
 LOCUS AX332462 419 bp DNA linear PAT 09-JAN-2002
 DEFINITION Sequence 2971 from Patent WO0194629.
 ACCESSION AX332462
 VERSION AX332462.1 GI:18123096
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Young.P.E., Augustus.M., Carter.K.C., Edner.R., Endress.G.,
 Horrigan.S., Soppet.D.R. and Weaver.Z.
 Cancer gene determination and therapeutic screening using signature
 gene sets
 Patent: WO 0194629-A 2971 13-DEC-2001;
 Avalon Pharmaceuticals (US)

REFERENCE
 AUTHORS Young.P.E., Augustus.M., Carter.K.C., Edner.R., Endress.G.,
 Horrigan.S., Soppet.D.R. and Weaver.Z.
 TITLE Cancer gene determination and therapeutic screening using signature
 JOURNAL Patent: WO 0194629-A 2971 13-DEC-2001;
 Avalon Pharmaceuticals (US)
 FEATURES
 source location/Qualifiers
 1..419
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 121 a 63 c 117 g 118 t
 ORIGIN

Query Match 35.3%; Score 316; DB 6; Length 419;
 Best Local Similarity 99.5%; Pred. No. 2,3e-170;
 Matches 416; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 478 aagcaaaacagatccagagcctgacagcaattctcaaaacatgctagcagaagc 537
 Db 419 AAGCCAAACAGCATTCAGAGGCTGCCAGCAATTTCTCAACAAATGTCAGCTAAGAAAGC 360
 QY 538 ttgtctgtccttctgttagagctctgagcgcacactcttcaattaaattctcagcca 597
 Db 359 TTTGCTCTCTGTTGTAGAGGCTGTGAGGCCACACATTTCAATTAACATTTCTCAGCCA 300
 QY 598 agaagacagtgagcagacactcagacacactctctctccacactcactccacatgta 657
 Db 299 AGAAGACATGAGACACACTTACCAGACACTCTTTCTCTCCACACTCCTCCACTGTGA 240
 QY 658 cccacccctaaatcattcagtgctctcaaaaagcatgttttcaagaatcattgttg 717
 Db 239 CCCACCCCTAAATTCATTCAGTGTCTCAAAAAGCATGTTTTCAAAGATCAATTTGTTTG 180
 QY 718 ttgtctctctagtgctccag 777
 Db 179 TTGCTCTCTCTGTTGTGTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
 QY 778 gcttagagcttaattacctaagatcagagaaactgtatgcttccctagctagtgatcatt 837
 Db 119 GCTTAGGCTTAATTAATCTCAAAAGATTCAGGAAACGTAGCTTCCAGTACTGATATTT 60
 QY 838 aaccttaaatgcaatcaggaagtagcaaaacagaaagtcataataatttttaaatgctc 895
 Db 59 AACCTTAATGCAATCAGGAAAGTAGCAAAAGTAGCAATTAATTTTAAATGTC 2

```

RESULT 9
AX062416 312 bp DNA linear PAT 24-JAN-2001
LOCUS Sequence 43 from Patent WO0100828.
DEFINITION AX062416
ACCESSION AX062416
VERSION AX062416.1 GI:12540291
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 312)
AUTHORS Wang, T., Bangur, C.S., Lodes, M.J., Fanger, G.R., Vedvick, T.S.,
Carter, D., Retter, M.W. and Mannion, J.
TITLES Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0100828-A 43 04-JAN-2001;
CORONA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..312
BASE COUNT 100 a 82 c 69 g 61 t
ORIGIN
Query Match 34.9%; Score 312; DB 6; Length 312;
Best Local Similarity 100.0%; Pred. No. 4,7e-168;
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 325 caggaaagcgcccaagaatgtgaatgcaaaagtgttcctgagagcccgagaagaaa 384
D 1 CAGGAAGCGCGCCCAAGATGTGAGTGCAGAAAGATTGTTCCAGAGAGCCCGAGAGAAA 60
QY 385 ttctgacagtgctggtggtgcgaagaagcagtgccctgtgatacattccaaggcaat 444
D 61 TTCAATGACAGTGTGTGGCTGCGCAAGAGAGCGCCCTGTGATTCATTTCAAGGGCAAT 120
QY 445 gtgaagaagaacagacccaagaagcaccagaagaagcacaagaagcattcagaagctgc 504
D 121 GTGAAGAAGAACAAGACCAAGAGCCACGACGAGAAAGCCAAAGCAATTCAGAGCTGC 180
QY 505 cagaatattctcaagaatgtcagcttaagaagcttgcctgtgcttgaagagctctga 564
D 181 CAGCAATTTCTCAAAACATGTACGCTAAGAAAGCTTTGCTGCTGTGAGAGCTCTGA 240
QY 565 gcgcccaactctccaattcaaacattctcagcagaagaagcagtgagacacaccagac 624
D 241 GCGGCCACCTCTCCAAATTTCACATTTCTCAGCCAGAGAGACAGACACCTACAGAC 300
QY 625 actctctctctc 636
D 301 ACTCTCTCTCTC 312
RESULT 10
E40584 291 bp DNA linear PAT 31-JAN-2002
LOCUS Novel protein and process for producing it.
DEFINITION E40584
ACCESSION E40584
VERSION E40584.1 GI:18625113
KEYWORDS JP 2000139479-A/4.
SOURCE Homo sapiens.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 291)
AUTHORS Ito, Y., Ogi, K., Tanaka, H. and Kitada, C.
TITLES Novel protein and process for producing it
JOURNAL Patent: JP 2000139479-A 4 23-MAY-2000;
TAKEDA CHEM IND LTD

```

```

COMMENT OS Homo sapiens (human)
PN JP 2000139479-A/4
PD 23-MAY-2000
PF 02-SEP-1999 JP 1999248414
PR
PI YASUAKI ITO, KAZUHIRO OGI, HIDEYUKI TANAKA, CHIERO KITADA
C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00, PC
A61K31/00,
PC A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
PC A61K38/00,
PC A61K45/00,A61K48/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,
PC G01N33/50//
CC A61K39/395,C12P21/08,(C12P21/02,C12R1:91),C12N15/00,A61K37/02
FH Key
FT source Location/Qualifiers
1..291
location/Qualifiers
1..291
BASE COUNT 90 a 72 c 79 g 50 t
ORIGIN
Query Match 32.5%; Score 291; DB 6; Length 291;
Best Local Similarity 100.0%; Pred. No. 6.1e-156;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 262 agcctgaatccaggggtcgcagagggccacagagggccagagcctcctagagatgg 321
D 1 AGCCTGAATCCAGGGGTGCGCCAGAGGCCACAGGAGCCAGAGGCTCTAGAGATGG 60
QY 322 ctccaggaagggcgccagaagaatgtgagtgcaaaagtgttcctgagagcccgagaaga 381
D 61 CTCAGAGAGCGCGCCAGAGATGTGAGTGCAGAAAGATTGTTCTGAGAGCCCGAGAGAA 120
QY 382 aaattcatgacgtgtctgtggtgcgtgcgaagaagcagtgccctgtgataatttcaaggcc 441
D 121 AAATTCATGACGTGTCTGTGGCTGCGCCAAAGAGACATGCTGCTGTGATTCATTTCAAGGGCA 180
QY 442 aatgtgaagaagaacagacccaagaagcaccagaagaagcacaagaagcattcagaagcc 501
D 181 AATGTGAAGAAGAACAAGACCAAGAGCCACGACGAGAAAGCCAAAGCAATTCAGAGGCC 240
QY 502 tgcgaagaattctctcaacaagtgcagctagaagaagcttgcctgccttg 552
D 241 TGCCAGCAATTTCTCAAAACATGTACGCTAAGAAAGCTTTGCTGCTGCTTTG 291
RESULT 11
AX182077 97 bp DNA linear PAT 06-AUG-2001
LOCUS Sequence 12 from Patent WO0142792.
DEFINITION AX182077
ACCESSION AX182077
VERSION AX182077.1 GI:15133350
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 97)
AUTHORS Schlegel, R., Deeds, J., Berger, A. and Zhao, X.
TITLES Diagnosis of cervical cancer using marker proteins
JOURNAL Patent: WO 0142792-A 12 14-JUN-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source Location/Qualifiers
1..97
location/Qualifiers
1..97
BASE COUNT 18 a 16 c 37 g 24 t 2 others
ORIGIN

```

Query Match 7.6%; Score 68; DB 6; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.8e-27;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 591 tcagcaagaagacagtgagcacactacacagacactctctccacactctcc 650
|||||
DB 91 TCAGCCAGAGACAGACGACACCTACACGACACTCTCTCCACCTCAGCTCC 32

QY 651 cactgtac 658
|||||
DB 31 CACTGTAC 24

RESULT 12
AX187025/c 456 bp DNA 1linear PAT 06-AUG-2001
LOCUS AX187025
DEFINITION Sequence 2720 from Patent WO0142467.
ACCESSION AX187025
VERSION AX187025.1 GI:15138468
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 456)
AUTHORS Schlegel, R., Deeds, J., Berger, A. and Zhao, X.
TITLE Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer
JOURNAL Patent: WO 0142467-A 2720 14-JUN-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source 1. 456
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 93 a 97 c 105 g 161 t
ORIGIN

Query Match 4.7%; Score 42; DB 6; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.6e-12;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 618 accagacactctctctccacactcactccacactgac 659
|||||
DB 42 ACCAGACACTCTCTCTCCACCTCAGCTCTCCACTGTACC 1

RESULT 13
AX185599/c 518 bp DNA 1linear PAT 06-AUG-2001
LOCUS AX185599
DEFINITION Sequence 1294 from Patent WO0142467.
ACCESSION AX185599
VERSION AX185599.1 GI:15137000
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 518)
AUTHORS Schlegel, R., Deeds, J., Berger, A. and Zhao, X.
TITLE Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer
JOURNAL Patent: WO 0142467-A 1294 14-JUN-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source 1. 518
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 96 a 101 c 104 g 160 t 57 others
ORIGIN

Query Match 4.7%; Score 42; DB 6; Length 518;
Best Local Similarity 100.0%; Pred. No. 2.6e-12;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 618 accagacactctctctccacactcactccacactgac 659
|||||
DB 64 ACCAGACACTCTCTCTCCACCTCAGCTCTCCACTGTACC 23

RESULT 14
E40591/c 51 bp DNA 1linear PAT 31-JAN-2002
LOCUS E40591
DEFINITION Novel protein and process for producing it.
ACCESSION E40591
VERSION E40591.1 GI:18625120
KEYWORDS JP 2000139479-A/11.
SOURCE JP 2000139479-A/11.
ORGANISM synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 51)
AUTHORS Ito, Y., Ogi, K., Tanaka, H. and Kitada, C.
TITLE Novel protein and process for producing it
JOURNAL Patent: JP 2000139479-A 11 23-MAY-2000;
TAKEDA CHEM IND LTD
COMMENT OS Artificial Sequence
PN JP 2000139479-A/11
PD 23-MAY-2000
PF 02-SEP-1999 JP 1999248414
PR

PI YASUAKI ITO, KAZUHIRO OGI, HIDEYUKI TANAKA, CHIEKO KITADA PC
C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00, PC
A61K31/00,
PC A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
PC A61K38/00,
PC A61K45/00,A61K48/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,
PC G01N33/50//
PC A61K39/395,C12P21/08,(C12P21/02,C12R1:91),C12N15/00,A61K37/02
CC
FH
FT key
FT source 1. 51
Location/Qualifiers

FEATURES
source 1. 51
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 14 a 13 c 11 g 13 t
ORIGIN

Query Match 4.5%; Score 40; DB 6; Length 51;
Best Local Similarity 100.0%; Pred. No. 4.1e-11;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 aaacaatgctcagctaaagaagcttgctcgtccttgtagg 556
|||||
DB 51 AAACAATGCTCAGCTAAGAAGCTTGTGCTGCTTTGTAGG 12

RESULT 15
AC011513 40520 bp DNA 1linear PRI 21-APR-2000
LOCUS AC011513
DEFINITION Homo sapiens chromosome 19 clone L1NF-112E5, complete sequence.
ACCESSION AC011513
VERSION AC011513.2 GI:7630362
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 40520)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 40520)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 40520)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Apr 21, 2000 this sequence version replaced gi:6015239.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
STS Content:
WI-7069 G06380.
FEATURES
source Location/Qualifiers
1..40520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="ILNLF-112E5"
BASE COUNT 11546 a 9580 c 8847 g 10547 t
ORIGIN

Query Match 4.18; Score 37; DB 9; Length 40520;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 239 tgcgtatgcatgtctctcagcagcctgaatccag 275
|||||
Db 10677 TcGTGATGTCATGCTCTCTAGCAGCCGATCCAGG 10713

Search completed: September 16, 2002, 06:32:07
Job time: 12405 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:31:41 ; Search time 390.66 Seconds
(without alignments)
3933.442 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895

Sequence: 1 ctaatctgtctacgacgc.....aataatattttaatgac 895

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 895 | 100.0 | 895 | 21 | AAZ29726 |
| 2 | 895 | 100.0 | 895 | 22 | AAH77949 |
| 3 | 895 | 100.0 | 1171 | 22 | AAH77951 |
| 4 | 858 | 95.9 | 870 | 21 | AAZ98202 |
| 5 | 852 | 95.2 | 870 | 21 | AAZ65001 |
| 6 | 852 | 95.2 | 870 | 22 | AAZ55994 |
| 7 | 852 | 95.2 | 870 | 22 | AAZ92070 |
| 8 | 852 | 95.2 | 870 | 22 | AAZ41417 |
| 9 | 807 | 90.2 | 1148 | 22 | ABA09257 |

| | | | | | | |
|----|-----|------|-------|----|----------|---------------------|
| 10 | 630 | 70.4 | 654 | 21 | AAC59829 | Human secreted pro |
| 11 | 593 | 66.3 | 849 | 21 | AAAO8343 | Human TGC-440 secr |
| 12 | 427 | 47.7 | 533 | 22 | AAZ68195 | Human lung tumour |
| 13 | 357 | 39.9 | 357 | 21 | AAAO8344 | Human TGC-440 secr |
| 14 | 335 | 37.4 | 365 | 20 | AAZ40454 | Human secreted pro |
| 15 | 312 | 34.9 | 312 | 22 | AAZ68125 | Human lung tumour |
| 16 | 307 | 34.3 | 511 | 19 | AAZ69618 | Human secreted pro |
| 17 | 291 | 32.5 | 291 | 21 | AAAO8345 | Mature human TGC-4 |
| 18 | 176 | 19.7 | 235 | 20 | AAZ40453 | Human secreted pro |
| 19 | 117 | 13.1 | 195 | 16 | AAZ25820 | Human gene signatu |
| 20 | 68 | 7.6 | 97 | 22 | AAZ07713 | Cervical cancer pr |
| 21 | 42 | 4.7 | 456 | 22 | AAH71445 | Human cervical can |
| 22 | 42 | 4.7 | 518 | 22 | AAH70020 | Human cervical can |
| 23 | 40 | 4.5 | 51 | 21 | AAAO8356 | Human TGC-440 secr |
| 24 | 31 | 3.5 | 47 | 21 | AAAO8355 | Human TGC-440 secr |
| 25 | 28 | 3.1 | 291 | 21 | AAAO8351 | Mature mouse TGC-4 |
| 26 | 28 | 3.1 | 357 | 21 | AAAO8350 | Mouse TGC-440 secr |
| 27 | 28 | 3.1 | 764 | 21 | AAAO8349 | Mouse TGC-440 secr |
| 28 | 26 | 2.9 | 40 | 21 | AAAO8353 | Human TGC-440 secr |
| 29 | 26 | 2.9 | 40 | 21 | AAAO8354 | Human TGC-440 secr |
| 30 | 22 | 2.5 | 2831 | 22 | AAZ03409 | Human reproductive |
| 31 | 22 | 2.5 | 33249 | 23 | ABZ15430 | Drosophila melanog |
| 32 | 20 | 2.2 | 20 | 21 | AAC59888 | Oligonucleotide pr |
| 33 | 20 | 2.2 | 20 | 22 | AAH77954 | PCR primer for CDN |
| 34 | 20 | 2.2 | 41 | 22 | AAZ12691 | Human SNP oligonuc |
| 35 | 20 | 2.2 | 467 | 22 | AAZ13198 | Human breast cancer |
| 36 | 20 | 2.2 | 469 | 22 | AAZ20057 | Human breast cancer |
| 37 | 20 | 2.2 | 481 | 22 | AAZ20068 | Human breast cancer |
| 38 | 20 | 2.2 | 641 | 22 | AAZ13187 | Human breast cancer |
| 39 | 20 | 2.2 | 3935 | 23 | ABZ16780 | Drosophila melanog |
| 40 | 20 | 2.2 | 7852 | 22 | ABZ14491 | Human nervous syst |
| 41 | 20 | 2.2 | 10518 | 23 | ABZ19372 | Drosophila melanog |
| 42 | 20 | 2.2 | 10616 | 22 | AAZ65565 | Human immune/haema |
| 43 | 20 | 2.2 | 10616 | 22 | AAZ65567 | Human immune/haema |
| 44 | 20 | 2.2 | 21732 | 22 | ABZ14492 | Human nervous syst |
| 45 | 20 | 2.2 | 21732 | 22 | AAZ61131 | Human immune/haema |

ALIGNMENTS

RESULT 1
AAZ29726
ID AAZ29726 standard; DNA: 895 BP.
XX
AC AAZ29726;
XX
DT 27-MAR-2000 (first entry)
XX
DE Human lung specific gene-3.
XX
KW Lung Specific Gene; LSG; human; diagnostic marker;
KW prognosticate; lung cancer; diagnosis; ds.
XX
OS Homo sapiens.
XX
PM WO9960160-A1.
XX
PD 25-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10344.
XX
PR 21-MAY-1998; 98US-0086212.
XX
PA (DIAD-) DIADEXUS LLC.
XX
PI Yang F, Macina RA, Sun Y;
XX
DR WPT: 2000-116320/10.
XX
PT A new method for diagnosing, monitoring and staging lung cancer -
XX
PS Example 1; Pages 35-36; 40pp; English.

XX The present sequence is a lung specific gene (LSG) from human
 CC clone ID 126263. The LSG has high level of tissue specificity for lungs
 CC and is overexpressed in cancerous tissues. The sequence serves as a
 CC diagnostic marker for detecting, monitoring, staging and prognosticating
 CC lung cancer. The diagnosis involves comparing levels of LSG in samples
 CC obtained from patient and normal control.

XX
 SQ Sequence 895 BP; 245 A; 250 C; 183 G; 217 T; 0 other;

Query Match 100.0%; Score 895; DB 21; Length 895;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaatcgttgcgttaacgaagacagcagctacccctggttgcgcctcaatgga 60
 Db 1 ctaatcgttgcgttaacgaagacagcagctacccctggttgcgcctcaatgga 60
 QY 61 acgctgcttggaactaaagcatagaccacagctgagtaacctgacatgacatccc 120
 Db 61 acgctgcttggaactaaagcatagaccacagctgagtaacctgacatgacatccc 120
 QY 121 caggatcagagagcctccagcaggaagacccctcattatattctcaagcaacttaagct 180
 Db 121 caggatcagagagcctccagcaggaagacccctcattatattctcaagcaacttaagct 180
 QY 181 gcaccgacagctgcatgaaagtcttaactctcctcctcctggttgcgcacatag 240
 Db 181 gcaccgacagctgcatgaaagtcttaactctcctcctcctggttgcgcacatag 240
 QY 241 ctgagtccatgctctcttaagcagcctgaatccagggctgcagagagccacagggacga 300
 Db 241 ctgagtccatgctctcttaagcagcctgaatccagggctgcagagagccacagggacga 300
 QY 301 ggcacagctctcaggaatgctccaggaaggcgccaaagatgtagtcaagaattgg 360
 Db 301 ggcacagctctcaggaatgctccaggaaggcgccaaagatgtagtcaagaattgg 360
 QY 361 ttctctgagagcccgagaaagaattcaatgacagtgctggctgcacaaagacagctgc 420
 Db 361 ttctctgagagcccgagaaagaattcaatgacagtgctggctgcacaaagacagctgc 420
 QY 421 cccctgcatcattccaaggcgaatgtagaagaagaacacacaaagacccacagaag 480
 Db 421 cccctgcatcattccaaggcgaatgtagaagaagaacacacaaagacccacagaag 480
 QY 481 ccaacaacagatccagagcctgcccagcaattctcaacaatgtagtaagaagctt 540
 Db 481 ccaacaacagatccagagcctgcccagcaattctcaacaatgtagtaagaagctt 540
 QY 541 gctctccttctgtagagcctgtagggccacactctccaattaaacattccagccaaga 600
 Db 541 gctctccttctgtagagcctgtagggccacactctccaattaaacattccagccaaga 600
 QY 601 agacagtgagcacactacacagacactctctctccacccctccactccacatgaccc 660
 Db 601 agacagtgagcacactacacagacactctctctccacccctccactccacatgaccc 660
 QY 661 acccctaataatcattcagtgctctcaaaaagcatgttttcaagatcatattgtttgtg 720
 Db 661 acccctaataatcattcagtgctctcaaaaagcatgttttcaagatcatattgtttgtg 720
 QY 721 cctctcctagtgctctctctctctctcagcttaagcctggtgcctcccttaaccagagct 780
 Db 721 cctctcctagtgctctctctctctctcagcttaagcctggtgcctcccttaaccagagct 780
 QY 781 taggcttaataacctgaagaatccaggaactgtagcttctaactagtgatcatttaac 840
 Db 781 taggcttaataacctgaagaatccaggaactgtagcttctaactagtgatcatttaac 840
 QY 841 cctaaatgcaatcaggaagtagcaacagagatcaataatttttaaatgtc 895
 Db 841 cctaaatgcaatcaggaagtagcaacagagatcaataatttttaaatgtc 895

Db 841 cctaaatgcaatcaggaagtagcaacagagatcaataatttttaaatgtc 895

RESULT 2

AAH77949

ID AAH77949 standard; DNA; 895 BP.

XX AAH77949;

AC 13-NOV-2001 (first entry)

DE Nucleotide sequence of a human Lng104 polypeptide.

KW Human; Lung cancer specific gene; LSG; Lng104; lung cancer; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 196..555

FT /tag= "a"

FT /product= "Lng104"

PN WO200161055-A2.

XX 23-AUG-2001.

XX 20-FEB-2001: 2001WO-US05674.

XX 17-FEB-2000: 2000US-0183188.

XX (DIAD-) DIADEXUS INC.

XX Chen S, Sun Y, Macina RA;

XX MPI: 2001-529917/58.

XX P-PSDB: AM63977.

XX New lung cancer specific gene for the treatment and diagnosis of lung cancer -

XX Claim 1; Page 113; 11pp; English.

CC The present sequence represents a human lung cancer specific gene (LSG), and encodes a polypeptide designated Lng104. LSGs are useful in the treatment and diagnosis of lung cancer. The treatment of lung cancer comprises the administration of a molecule which down regulates the expression of an LSG. An immune response can be mounted against a target cell expressing an LSG. Identification of potential therapeutic agents for use in imaging and treating lung cancer which comprises screening molecules for an ability to bind to or decrease expression of an LSG relative to LSG in the absence of the agent where the ability of a molecule to bind to the LSG or decrease expression of the LSG is indicative of the molecule being useful in imaging and treating lung cancer.

XX
 SQ Sequence 895 BP; 245 A; 250 C; 183 G; 217 T; 0 other;

Query Match 100.0%; Score 895; DB 22; Length 895;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaatcgttgcgttaacgaagacagcagctacccctggttgcgcctcaatgga 60
 Db 1 ctaatcgttgcgttaacgaagacagcagctacccctggttgcgcctcaatgga 60
 QY 61 acgctgcttggaactaaagcatagaccacagctgagtaacctgacatgacatccc 120
 Db 61 acgctgcttggaactaaagcatagaccacagctgagtaacctgacatgacatccc 120
 QY 121 caggatcagagagcctccagcaggaagacccctcattatattctcaagcaacttaagct 180
 Db 121 caggatcagagagcctccagcaggaagacccctcattatattctcaagcaacttaagct 180

QY 181 gcaccgacgttcgatagaagttcctaattcttcctcctctgtgtgtccactaatg 240
|||||
Db 181 gcaccgacggttcgatagaagttcctaattcttcctcctctgtgtgtccactaatg 240
QY 241 ctgatttcacatggtctcttagcagcctgtaalccaggggttcgacagagccacagagcga 300
|||||
Db 241 ctgatttcacatggtctcttagcagcctgtaalccaggggttcgacagagccacagagcga 300
QY 301 ggcacaggtcttagagatggtccaggaagggcgcaagaatgtagtgcaagaatttg 360
|||||
Db 301 ggcacaggtcttagagatggtccaggaagggcgcaagaatgtagtgcaagaatttg 360
QY 361 ttctctagagcccgagaaataatcatgacagtgtcttggtgtccaaagaagcagtgc 420
|||||
Db 361 ttctctagagcccgagaaataatcatgacagtgtcttggtgtccaaagaagcagtgc 420
QY 421 ccctgtgatcatctcaagggtcaatgttgaagaacaaacagaccacaaaggccacagaaag 480
|||||
Db 421 ccctgtgatcatctcaagggtcaatgttgaagaacaaacagaccacaaaggccacagaaag 480
QY 481 ccaaaagaagcatccagagcctgcagcaattctcaaaatgtcagctaaagaagcttt 540
|||||
Db 481 ccaaaagaagcatccagagcctgcagcaattctcaaaatgtcagctaaagaagcttt 540
QY 541 gctctgcttcttagagagctctgagcgccacactctccaattaaacattctcagcaga 600
|||||
Db 541 gctctgcttcttagagagctctgagcgccacactctccaattaaacattctcagcaga 600
QY 601 agacagctgagcacacacacacagacactctctctccacactccacactccacactgaccc 660
|||||
Db 601 agacagctgagcacacacacacacagacactctctctccacactccacactccacactgaccc 660
QY 661 acccctaataatcattccagtgctctcaaaaagcagtcttccaagatcatttgctgtg 720
|||||
Db 661 acccctaataatcattccagtgctctcaaaaagcagtcttccaagatcatttgctgtg 720
QY 721 cctctctagtgctctctctctctcgtcagctcttagagcttgccctccctaccagagct 780
|||||
Db 721 cctctctagtgctctctctctctcgtcagctcttagagcttgccctccctaccagagct 780
QY 781 taggcttaattcactgtaaaagatccaggaactgtagcttccagtagtgatgtcattaac 840
|||||
Db 781 taggcttaattcactgtaaaagatccaggaactgtagcttccagtagtgatgtcattaac 840
QY 841 cttaaatgcaatcaggaagtagcaaacagagatcaataatttttaattgctc 895
|||||
Db 841 cttaaatgcaatcaggaagtagcaaacagagatcaataatttttaattgctc 895

RESULT 3
AAH77951
ID AAH77951 standard; DNA; 1171 BP.
AC AAH77951;
XX
DT 13-NOV-2001 (first entry)
XX
DE Nucleotide sequence of a human Lng104 polypeptide.
XX
KW Human; lung cancer specific gene; LSG; Lng104; lung cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 217..576
FT /*tag= a
FT /product= "Lng104"
XX
PN WO200161055-A2.
XX
PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05674.
XX
XX 17-FEB-2000; 2000US-0183188.
XX
PA (DIAD-) DIADEXUS INC.
XX
XX Chen S, Sun Y, Macina RA;
XX
XX MPI: 2001-529917/58.
DR P-PSDB; AAC63977.
PT
PT New lung cancer specific gene for the treatment and diagnosis of lung
PT cancer -
PS
PS Claim 1; Page 117; 119pp; English.
XX
XX The present sequence represents a human lung cancer specific gene
CC (LSG), and encodes a polypeptide designated Lng104. LSGs are useful
CC in the treatment and diagnosis of lung cancer. The treatment of lung
CC cancer comprises the administration of a molecule which down regulates
CC the expression of an LSG. An immune response can be mounted against a
CC target cell expressing an LSG. Identification of potential therapeutic
CC agents for use in imaging and treating lung cancer which comprises
CC screening molecules for an ability to bind to or decrease expression
CC of an LSG relative to LSG in the absence of the agent where the ability
CC of a molecule to bind to the LSG or decrease expression of the LSG is
CC indicative of the molecule being useful in imaging and treating lung
CC cancer.
XX
XX Sequence 1171 BP; 333 A; 320 C; 225 G; 293 T; 0 other;

Query Match 100.0%; Score 895; DB 22; Length 1171;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctatctgttagcttaacagacagacagcgtcacctacactgcttctgcgcctcaaatggga 60
Db 22 ctatctgttagcttaacagacagacagcgtcacctacactgcttctgcgcctcaaatggga 81
QY 61 acgctgcttgggactaaagcatalagaccacagcgtgagatccctgagctgacatccc 120
Db 82 acgctgcttgggactaaagcatalagaccacagcgtgagatccctgagctgacatccc 141
QY 121 caaggatcaggaagcctccagcaggaaccttcattatattcttcaagcaacttaagct 180
Db 142 caaggatcaggaagcctccagcaggaaccttcattatattcttcaagcaacttaagct 201
QY 181 gcaccgacgttcgatagaagttcctaattcttcctcctctgtgtgtccactaatg 240
Db 202 gcaccgacggttcgatagaagttcctaattcttcctcctctgtgtgtccactaatg 261
QY 241 ctgatttcacatggtctcttagcagcctgtaalccaggggttcgacagagccacagagcga 300
Db 262 ctgatttcacatggtctcttagcagcctgtaalccaggggttcgacagagccacagagcga 321
QY 301 ggcacaggtcttagagatggtccaggaagggcgcaagaatgtagtgcaagaatttg 360
Db 322 ggcacaggtcttagagatggtccaggaagggcgcaagaatgtagtgcaagaatttg 381
QY 361 ttctctagagcccgagaaataatcatgacagtgtcttggtgtccaaagaagcagtgc 420
Db 382 ttctctagagcccgagaaataatcatgacagtgtcttggtgtccaaagaagcagtgc 441
QY 421 ccctgtgatcatctcaagggtcaatgttgaagaacaaacagaccacaaaggccacagaaag 480
Db 442 ccctgtgatcatctcaagggtcaatgttgaagaacaaacagaccacaaaggccacagaaag 501
QY 481 ccaaaagaagcatccagagcctgcagcaattctcaaaatgtcagctaaagaagcttt 540
Db 502 ccaaaagaagcatccagagcctgcagcaattctcaaaatgtcagctaaagaagcttt 561
QY 541 gctctgcttcttagagagctctgagcgccacactctccaattaaacattctcagcaga 600

| | | |
|----|-------------|---------------|
| PR | 12-JUN-1996 | 98US-00891105 |
| PR | 16-JUN-1996 | 98US-00895440 |
| PR | 16-JUN-1996 | 98US-00895512 |
| PR | 16-JUN-1996 | 98US-00895514 |
| PR | 17-JUN-1996 | 98US-00895532 |
| PR | 17-JUN-1996 | 98US-00895558 |
| PR | 17-JUN-1996 | 98US-00895598 |
| PR | 17-JUN-1996 | 98US-00895599 |
| PR | 17-JUN-1996 | 98US-00896000 |
| PR | 17-JUN-1996 | 98US-00896533 |
| PR | 18-JUN-1996 | 98US-00898081 |
| PR | 18-JUN-1996 | 98US-00899077 |
| PR | 18-JUN-1996 | 98US-00899088 |
| PR | 19-JUN-1996 | 98US-00899478 |
| PR | 19-JUN-1996 | 98US-00899488 |
| PR | 19-JUN-1996 | 98US-00899522 |
| PR | 22-JUN-1996 | 98US-00902466 |

PR 23-JU

| | | |
|----|--------------|--------------|
| PR | 23-JUN-1998; | 98US-0090355 |
| PR | 24-JUN-1998; | 98US-0090429 |

PR 24-JU

PR 24-JU

PR 24-JU

| | |
|----|---------|
| PR | 24 - JU |
| PR | 24 - JU |

PR 25-JU

PR 25-JU

PR 25-JU

PR 25-JU

PR 26-JU

PR 01-JU

PR 01-JU

PR 02-JU

PR 02-JU

PR 02-JU

PR 02-JU

PR 07-JU

PR 10-JU

PR 30-JU

PR 04-AU

PR 04-AU

PR 04-AU

PR 10-AU

PR 10-AU

PR 11-AU

PR 17-AU

```
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
XX (GETH ) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
DR WPI: 2000-072883/06.
DR P-PSDB; AAY66668.
XX
PT Membrane-bound proteins and related nucleotide sequences -
XX
XX Claim 2; Fig 98; 822pp; English.
XX
PS The invention provides membrane-bound PRO polypeptides and
XX CC polynucleotides encoding them. The PRO sequences of the invention were
XX CC identified based on extracellular domain homology screening. The PRO
XX CC sequences have homology with proteins including LDL receptors, TIE
XX CC ligands and various enzymes. The membrane-bound proteins and receptor
XX CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
XX CC immunoadhesins, for instance, can be used as therapeutic agents to block
XX CC receptor-ligand interactions. The membrane-bound proteins can also be
XX CC employed for screening of potential peptide or small molecule inhibitors
XX CC of the relevant receptor/ligand interaction. The PRO encoding sequences
XX CC are useful as hybridization probes, in chromosome and gene mapping and in
XX CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
XX CC will also be useful for the preparation of PRO polypeptides, especially
XX CC by recombinant techniques.
XX
SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;
```

```
Query Match 95.2%; Score 852; DB 21; Length 870;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 44 ctgcgccttaattggaagctgcgcctggagtaagaatagacacacagcgtgagatc 103
Db 1 cctgcgccttaattggaagctgcgcctggagtaagaatagacacacagcgtgagatc 60
QY 104 ctgacctgagatcatccacaggaatcagagcctcagacaggaaggaaccttcattatctc 163
Db 61 ctgacctgagatcatccacaggaatcagagcctcagacaggaaggaaccttcattatctc 120
```

```
QY 164 tcaagcaactatcacagctgcacccagcaagcttgcatgaaagtcttaattcttcctctcc 223
Db 121 tcaagcaactatcacagctgcacccagcaagcttgcatgaaagtcttaattcttcctctcc 180
QY 224 tgttcgcgcactaatgctgatgtccatgtgtctccttaagcagctgaatcccggtgtgcc 293
Db 181 tgttcgcgcactaatgctgatgtccatgtgtctccttaagcagctgaatcccggtgtgcc 240
QY 284 gaggccacaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 343
Db 241 gaggccacaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
QY 344 gtgagtgcagaagattgttctctgaagcccgaggaagaaattcagacagtgctcgggc 403
Db 301 gtgagtgcagaagattgttctctgaagcccgaggaagaaattcagacagtgctcgggc 360
QY 404 tgcacaagaagcagtgcccttgatcatcattcagaagcagtggaagaacaagaacacc 463
Db 361 tgcacaagaagcagtgcccttgatcatcattcagaagcagtggaagaacaagaacacc 420
QY 464 aaagcaccacagaagaagcacaacagcattccagagcctgcagcaattctcacaacat 523
Db 421 aaagcaccacagaagaagcacaacagcattccagagcctgcagcaattctcacaacat 480
QY 524 gtacgttaagaagcttgctcctcttgtagagctctgagcgccacattctccaatta 583
Db 481 gtacgttaagaagcttgctcctcttgtagagctctgagcgccacattctccaatta 540
QY 584 aacattctcagccacaagaagcagtgagcacacactcacaacactctctccacacc 643
Db 541 aacattctcagccacaagaagcagtgagcacacactcacaacactctctccacacc 600
QY 644 actctccacgttaccaccccttaataatcattccagtgctctcacaagaatgttttcaa 703
Db 601 actctccacgttaccaccccttaataatcattccagtgctctcacaagaatgttttcaa 660
QY 704 gatcatttggttgctcctctcagtgctctctctctctctctctctctctctctctctct 763
Db 661 gatcatttggttgctcctctcagtgctctctctctctctctctctctctctctctctct 720
QY 764 ctcccttaaccagcgttagccttaattacactgaagaattccagaagaactgtagcttcc 823
Db 721 ctcccttaaccagcgttagccttaattacactgaagaattccagaagaactgtagcttcc 780
QY 824 agctagtgcatttaaccttaattgaatgcaggaagagagaagaagaagatcaataata 883
Db 781 agctagtgcatttaaccttaattgaatgcaggaagagagaagaagaagatcaataata 840
QY 884 ttttaaatgctc 895
Db 841 ttttaaatgctc 852
```

```
RESULT 6
AAS4594
ID AAS4594 standard; cDNA: 870 BP.
```

```
XX AAS4594:
```

```
DT 18-DEC-2001 (first entry)
```

```
DE Human DNA encoding PRO polypeptide sequence #70.
```

```
XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
XX PCR primer.
```

```
XX Homo sapiens.
```

```
OS WO200168848-A2.
```

```
XX
```

PD 20-SEP-2001.
XX
XX 28-FEB-2001; 2001MO-US06520.
PR 01-MAR-2000; 2000MO-US05601.
PR 02-MAR-2000; 2000MO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000MO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191027P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192555P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000MO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 03-MAY-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000MO-US13705.
PR 22-MAY-2000; 2000MO-US14042.
PR 30-MAY-2000; 2000MO-US14941.
PR 02-JUN-2000; 2000MO-US15264.
PR 02-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000MO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000MO-US23328.
PR 08-NOV-2000; 2000MO-US30952.
PR 01-DEC-2000; 2000MO-US32678.
PR 20-DEC-2000; 2000MO-US34956.
XX
XX (GETH) GENENTECH INC.
XX
PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PI, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI; 2001-602746/68.
DR P-PSDB: AAU29093.
XX
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
XX Claim 2; Fig 139; 774pp; English.
XX
XX Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC primers for PRO polypeptides of the invention. The sequences of the
CC invention can be used to detect the presence of a tumour in a mammal by
CC comparing the level of expression of a PRO polypeptide in a test sample
CC of cells from the animal and a control sample of normal cells, whereby a
CC higher level of expression in the test sample indicates the presence of a
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC pigs, goats and rabbits but are preferably human. The polypeptides can be
CC used to stimulate tumour necrosis factor (TNF) alpha release from human
CC blood, when contacted with it. A specific polypeptide can be used to
CC stimulate the proliferation or differentiation of chondrocyte cells. The
CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids

CC can be used for genetic analysis of individuals with genetic disorders.
XX
XX Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;
SQ

Query Match 95.2%; Score 852; DB 22; Length 870;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgccctaataatggaacgctgagctggaactaaagcataaccacagcgtgagatc 103
DB 1 ctgcgccctaataatggaacgctgagctggaactaaagcataaccacagcgtgagatc 60
QY 104 ctgacgtgagatctccacagggatccagagcctccacagcgggaaccccttcattatct 163
DB 61 ctgacgtgagatctccacagggatccagagcctccacagcgggaaccccttcattatct 120
QY 164 tcaaggaacttaacgctgacccagcagctgagatgataagttctatctctccctcc 223
DB 121 tcaaggaacttaacgctgacccagcagctgagatgataagttctatctctccctcc 180
QY 224 tcttctgccaactaatgctgatagtlccatagttctctagcagcctgaaatccagggtcgca 283
DB 181 tcttctgccaactaatgctgatagtlccatagttctctagcagcctgaaatccagggtcgca 240
QY 284 gaggccacagggagccgagccagcgtctctaggaagatgctccagggagccgccaagaat 343
DB 241 gaggccacagggagccgagccagcgtctctaggaagatgctccagggagccgccaagaat 300
QY 344 gtgagtgccaagaatltgtctctgagagcccgagaaataatccaagacgtgtctgggc 403
DB 301 gtgagtgccaagaatltgtctctgagagcccgagaaataatccaagacgtgtctgggc 360
QY 404 tgcacaagaagcagtgccctgtgatacttccaaggcaatgtgaagaacaagaacacc 463
DB 361 tgcacaagaagcagtgccctgtgatacttccaaggcaatgtgaagaacaagaacacc 420
QY 464 aagagccacacgaagaagcacaacaagaatctccagaaacccgccaagattctcaacaat 523
DB 421 aagagccacacgaagaagcacaacaagaatctccagaaacccgccaagattctcaacaat 480
QY 524 gtcagcctaagaagcttgtctctgctcttgaagagctctgaagcgcctccactctcaatta 583
DB 481 gtcagcctaagaagcttgtctctgctcttgaagagctctgaagcgcctccactctcaatta 540
QY 584 aacattctcagccaagaagaacagtgagacacactaccacaacactctctctccaccctc 643
DB 541 aacattctcagccaagaagaacagtgagacacactaccacaacactctctctccaccctc 600
QY 644 actctccacttaaccacccccaataatccaatccaagtcgtcctcaaaaagcatgttttcaa 703
DB 601 actctccacttaaccacccccaataatccaatccaagtcgtcctcaaaaagcatgttttcaa 660
QY 704 gatcaatttgttgtgtctctctctagtgctctctctctctctctctctctctctctctct 763
DB 661 gatcaatttgttgtgtct 720
QY 764 ctcccttaaccagcgttgcttaattacctcgaagaagattccaagaactgtagcttct 823
DB 721 ctcccttaaccagcgttgcttaattacctcgaagaagattccaagaactgtagcttct 780
QY 824 agctagtgcatcttaacctaataatgcaatcaggaagatgagcaaacaggaagtcataata 883
DB 781 agctagtgcatcttaacctaataatgcaatcaggaagatgagcaaacaggaagtcataata 840
QY 884 ttcttaacgtc 895
DB 841 ttcttaacgtc 852

RESULT 7
AAF92070
ID AAF92070 standard; cDNA; 870 BP.

```

XX AC AAF92070;
XX XX
XX 15-MAY-2001 (first entry)
XX DE Human PRO842 cDNA.
XX KW Human; PRO protein; mapping; ss.
XX OS Homo sapiens.
XX PN WO200116318-A2.
XX PD 08-MAR-2001.
XX PF 24-AUG-2000; 2000WO-US23328.
XX PR 01-SEP-1999; 99WO-US20111.
XX PR 15-SEP-1999; 99WO-US21080.
XX PR 07-DEC-1999; 99US-0169495.
XX PR 09-DEC-1999; 99US-0170262.
XX PR 11-JAN-2000; 2000US-0175481.
XX PR 18-FEB-2000; 2000WO-US04341.
XX PR 18-FEB-2000; 2000WO-US04342.
XX PR 22-FEB-2000; 2000WO-US04414.
XX PR 01-MAR-2000; 2000WO-US05601.
XX PR 03-MAR-2000; 2000US-0187202.
XX PR 23-APR-2000; 2000US-0193937.
XX PR 22-MAY-2000; 2000WO-US14042.
XX PR 05-JUN-2000; 2000US-0209832.
XX PA (GETH ) GENENTECH INC.
XX PI Epton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX DR WPI; 2001-183260/18.
XX DR P-PSDB; AAB87538.
XX PT Eighty four nucleic acids encoding PRO polypeptides, useful in
XX PT molecular biology, including use as hybridization probes, and in
XX PT chromosome and gene mapping.
XX PS Claim 2; Fig 25; 278pp; English.
XX CC The present sequence is the coding sequence for a human PRO polypeptide
XX CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO
XX CC antagonists or anti-PRO antibodies are useful for preparation of a
XX CC medicament useful in the treatment of a condition which is responsive to
XX CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO
XX CC protein may also be employed as molecular weight markers for protein
XX CC electrophoresis. The PRO coding sequence has applications in molecular
XX CC biology, including use as hybridisation probes, and in chromosome and
XX CC gene mapping.
XX SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

```

Query Match 95.2%; Score 852; DB 22; Length 870;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 44 ctgcgcctcaaatggaagcgtgctgagtaagaacatagaccacggctgagatc 103
DB 1 ctgcgcctcaaatggaagcgtgctgagtaagaacatagaccacggctgagatc 60
OY 104 ctgacctgagtcattcccaaggatcagaagctccacagggaggaaccttcattatct 163
DB 61 ctgacctgagtcattcccaaggatcagaagctccacagggaggaaccttcattatct 120
OY 164 tcaagaaacttaacgactgacagacagtgatgaagtcttattctcttccctctcc 223
DB 121 tcaagaaacttaacgactgacagacagtgatgaagtcttattctcttccctctcc 180

```

```

OY 224 tgtgtcgtccactaatgctgattgcatlccatggtctctctagcaagcctlgatccagggtctgcga 283
DB 181 tgtgtcgtccactaatgctgattgcatlccatggtctctctagcaagcctlgatccagggtctgcga 240
OY 284 gaggccacagggaccagggccaggctctctagggatggtctccagggaagcgcccaagaat 343
DB 241 gaggccacagggaccagggccaggctctctagggatggtctccagggaagcgcccaagaat 300
OY 344 gtgagtcaagaagtgttctctgagagcccggaagaataatctcatgacgtgtctgggc 403
DB 301 gtgagtcaagaagtgttctctgagagcccggaagaataatctcatgacgtgtctgggc 360
OY 404 tgcctaaagaagcagtgccctctgcatcattcaagggaatgtgaaagaacaagacacc 463
DB 361 tgcctaaagaagcagtgccctctgcatcattcaagggaatgtgaaagaacaagacacc 420
OY 464 aaagccacacagaaagacacaaacaaacatccagacgtctcagcaatttctcaacaat 523
DB 421 aaagccacacagaaagacacaaacaaacatccagacgtctcagcaatttctcaacaat 480
OY 524 gtcaagctaaagaagcttgcctctgctcttgaaggagctctgagccacactctccaat 583
DB 481 gtcaagctaaagaagcttgcctctgctcttgaaggagctctgagccacactctccaat 540
OY 584 aacattctcagccaagaagaagcagtgagcaacactaccagacatcttcttccaccc 643
DB 541 aacattctcagccaagaagaagcagtgagcaacactaccagacatcttcttccaccc 600
OY 644 actctccactgtaccacacccctaaatcattccagtgctctcaaaaagacatgtttcaa 703
DB 601 actctccactgtaccacacccctaaatcattccagtgctctcaaaaagacatgtttcaa 660
OY 704 gatcattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 763
DB 661 gatcattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
OY 764 ctcccttaaccagcgttagcttaattacctggaagatccaggaattgttagcttct 823
DB 721 ctcccttaaccagcgttagcttaattacctggaagatccaggaattgttagcttct 780
OY 824 agctagtgctattcaacttaaatgcaatcaggaagaatgacaaacagaaatcaata 883
DB 781 agctagtgctattcaacttaaatgcaatcaggaagaatgacaaacagaaatcaata 840
OY 884 tttttaaatgtc 895
DB 841 tttttaaatgtc 852

```

RESULT 8
 AAF44147
 ID AAF44147 standard; CDNA; 870 BP.
 AC AAF44147;
 XX
 XX 02-APR-2001 (first entry)
 DT
 XX
 XX Human PRO842 (UNQ473) nucleotide sequence SEQ ID NO:164.
 DE
 XX Human; secreted and transmembrane protein; PRO; cytosolic;
 KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
 KW diagnostic assay; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200073454-A1.
 PN
 XX
 XX 07-DEC-2000.
 PD
 XX
 XX 30-MAR-2000; 2000WO-US08439.
 PF
 XX
 XX 02-JUN-1999; 99WO-US12252.
 PR

PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 17-AUG-1999; 99US-0149396.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 08-OCT-1999; 99US-0158663.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
XX
XX (GETH) GENENTECH INC.
PA
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrarini N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Kijavini IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
XX WPI: 2001-032160/04.
DR P-PSDB: AAB65191.
XX
XX PRO polynucleotides used to produce polypeptides used to target
PT bioactive molecules such as toxins, radiolabels or antibodies, to
PT specific cells, to cause targeted cell death -
XX
XX
PS Claim 2; Fig 98; 935pp; English.
XX
XX The present invention describes human secreted and transmembrane PRO
CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC can be used for targeted delivery of bioactive molecules, such as
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC sequences, and their fragments, can be used as hybridisation probes, in
CC chromosomal and gene mapping, and in the generation of anti-sense RNA
CC and DNA. They may also be used to produce transgenic animals which are
CC used to develop and screen therapeutically useful reagents. The PRO
CC nucleotide and protein sequence can be used for tissue typing and in
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX
XX Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 22; Length 870;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgcctcaaatgggaacgtgcttgctggactaagaacacacacagctgagtc 103
DB 1 ctgcgcctcaaatgggaacgtgcttgctggactaagaacacacacagctgagtc 60
QY 104 ctgacctgagtcacccagggatcagagcctcagacgggaaccttcattatctt 163
DB 61 ctgacctgagtcacccagggatcagagcctcagacgggaaccttcattatctt 120
QY 164 tcaagcaacttaacagctcagccagacagtgatgaatcattccttcctccctcc 223

DB 121 tcaagcaacttaacagctcagccagacagtgatgaatcattccttcctccctcc 180
QY 224 tcttctgctccactaatgctcgtatcattcgtctctcagcagccttgatccagggtcgcca 283
DB 181 tcttctgctccactaatgctcgtatcattcgtctctcagcagccttgatccagggtcgcca 240
QY 284 gaggccacagggagcagcagcagccttcagagagatgctcaggaagacggcagcaaat 343
DB 241 gaggccacagggagcagcagcagccttcagagagatgctcaggaagacggcagcaaat 300
QY 344 gtgagtgcaagaatgtgtctcctgagagcccgagaagaatactcagacgtgtctgggc 403
DB 301 gtgagtgcaagaatgtgtctcctgagagcccgagaagaatactcagacgtgtctgggc 360
QY 404 tgcgaagaagagagtgccctgtgtaacttttaaggcagatgtggaagaaacaaagaccc 463
DB 361 tgcgaagaagagagtgccctgtgtaacttttaaggcagatgtggaagaaacaaagaccc 420
QY 464 aagagcaccacagaaagcacaacagcattccagagcctgcagcaatctcacaacat 523
DB 421 aagagcaccacagaaagcacaacagcattccagagcctgcagcaatctcacaacat 480
QY 524 gtcaagtaagaagcttgcctgtccttgtagaagctctgagcgccactctccaatla 583
DB 481 gtcaagtaagaagcttgcctgtccttgtagaagctctgagcgccactctccaatla 540
QY 584 aacattctcagcagaagaagacagtgagcaccctccagacactctctccccaactc 643
DB 541 aacattctcagcagaagaagacagtgagcaccctccagacactctctccccaactc 600
QY 644 actctccacatgaccacccccaataatcattcagtgctctcaaaaagatgttttcaa 703
DB 601 actctccacatgaccacccccaataatcattcagtgctctcaaaaagatgttttcaa 660
QY 704 gatcattgttgtgtctctcctagtgctctctctctctctcagctctagcctgtgccc 763
DB 661 gatcattgttgtgtctctcctagtgctctctctctctctcagctctagcctgtgccc 720
QY 764 ctcccttaaccagcttagcttaattaccctggaagaatctcaggaagaatgttagcttct 823
DB 721 ctcccttaaccagcttagcttaattaccctggaagaatctcaggaagaatgttagcttct 780
QY 824 agctatgtcatttaaccttaaatgcattcagacaggaagatgacaagaagatcaataa 883
DB 781 agctatgtcatttaaccttaaatgcattcagacaggaagatgacaagaagatcaataa 840
QY 884 ttcttaaatgtc 895
DB 841 ttcttaaatgtc 852

RESULT 9
ABA09257
ID ABA09257 standard; cDNA; 1148 BP.
XX
AC ABA09257;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human TGC-440 homologue-encoding cDNA, SEQ ID NO:1033.
XX
XX Human: cytokine; cell proliferation; tissue growth; immunomodulator; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;

KW antifungal; vulnery; antiulcer; ss.
XX OS. Homo sapiens.
XX PN WO200157188-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US03800.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX PA (HXSE-) HXSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-457740/49.
DR P-PSDB: ABB12013.
XX
XX Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
XX Claim 1; Page 867-868; 1963pp; English.

CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, hematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention.
XX
XX Sequence 1148 BP; 328 A; 309 C; 222 G; 289 T; 0 other;

Query Match 90.2%; Score 807; DB 22; Length 1148;
Best Local Similarity 99.9%; Pred. No. 0;

Matches 857; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 38 cctgtcttcgcctcaaatggagacgtgctgagcttaagacatagaccacagctg 97
|||||

Db 23 cctgtcttcgcctcaaatggagacgtgctgagcttaagacatagaccacagctg 82
QY 98 agtaccctgacctgagctatcccaaggatcaagagctccagcaggaaccttccatta 157
Db 83 agtaccctgacctgagctatcccaaggatcaagagctccagcaggaaccttccatta 142
QY 158 tatctctcaagcaactatacagctgcacccagagcttgcagatgaagtcttaactctcc 217
Db 143 tatctctcaagcaactatacagctgcacccagagcttgcagatgaagtcttaactctcc 202
QY 218 tctctcctgtctgcacactatgtcatgtctgtcttcaagctccagctccaggg 277
Db 203 tctctcctgtctgcacactatgtcatgtctgtcttcaagctccagctccaggg 262
QY 278 tcgccaagagccacaggaagccagagcttctgaagatggctccaggaagcgcc 337
Db 263 tcgccaagagccacaggaagccagagcttctgaagatggctccaggaagcgcc 322
QY 338 aagaatgtgagtgcaaggttggtcttgaagccccggagaagaataatcatgacagt 397
Db 323 aagaatgtgagtgcaaggttggtcttgaagccccggagaagaataatcatgacagt 382
QY 398 ctgggtgcacaagaagcagtgccctgtgacatcttcaaggcaatggaagaacaa 457
Db 383 ctgggtgcacaagaagcagtgccctgtgacatcttcaaggcaatggaagaacaa 442
QY 458 gaacccaaggaagccacagaaagccaaacagatctccagagctgcagcaatttcca 517
Db 443 gaacccaaggaagccacagaaagccaaacagatctccagagctgcagcaatttcca 502
QY 518 aacaaatgtaagtaagaagcttgcctgtcttgaagagcttgagccgacctctc 577
Db 503 aacaaatgtaagtaagaagcttgcctgtcttgaagagcttgagccgacctctc 562
QY 578 caattaacaatctcaagccaagaagcagtgagcacacctaccagacactctctcc 637
Db 563 caattaacaatctcaagccaagaagcagtgagcacacctaccagacactctctcc 622
QY 638 caactcacctccacactgacacccaccttaacttccagtgccttaaaaaagatgtt 697
Db 623 caactcacctccacactgacacccaccttaacttccagtgccttaaaaaagatgtt 682
QY 698 ttccaagatcattgttgtgtctctctctagtgctctctctctctctctctctctc 757
Db 683 ttccaagatcattgttgtgtctctctctctagtgctctctctctctctctctctc 742
QY 758 tgtgcccctccctaccacagcttaggcttaattacctgaaagattccaggaactgtag 817
Db 743 tgtgcccctccctaccacagcttaggcttaattacctgaaagattccaggaactgtag 802
QY 818 ctctctagctagtgcatcttaaccttaattcaatcaagaaagtagcaaacagatcaa 877
Db 803 ctctctagctagtgcatcttaaccttaattcaatcaagaaagtagcaaacagatcaa 862
QY 878 taatatattttaaatgtc 895
Db 863 taatatattttaaatgtc 880

RESULT 10
ID AAC59829 standard; DNA; 654 BP.
XX
AC AAC59829;
XX
DT 26-JAN-2001 (first entry)

XX Human secreted protein encoding DNA clone vq8 1.

XX Secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;
KW systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;
KW haemotopiasis regulation; tissue regrowth; wound healing; haemophilia;
KW Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;

KW contraceptive; infection; growth inhibition; hyperproliferative disorder;
KW psoriasis; ds.
XX
XX OS Homo sapiens.
XX WO200055375-A1.
XX
XX 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-US07285.
XX
XX 17-MAR-1999; 99US-0124808.
XX 17-MAR-1999; 99US-0124916.
XX 17-AUG-1999; 99US-0149639.
XX 01-OCT-1999; 99US-0157247.
XX 29-NOV-1999; 99US-0167824.
XX 15-FEB-2000; 2000US-0182711.
XX
XX (ALPH-) ALPHAGEN INC.
XX
XX Valenzuela D, Yuan O, Hoffman H, Hall J, Kaplejsko P;
PI WPI; 2000-638211/61.
XX P-PSDB; AAB34728.
XX
XX Novel proteins and polypeptides useful for the treatment of e.g
PT multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
PT cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and
PT ulcers .
XX
XX Claim 92; Page 441; 493pp; English.
XX
XX This invention relates to 59 human secreted proteins and the nucleotide
XX sequences encoding them. Sequences AAC59788-C59846 and AAB34687-B34745
XX represent the proteins and their encoding nucleotide sequences, and
XX sequences AAB34746-B34771 represent fragments of the proteins. Probes
XX for the DNA sequences are represented by sequences AAC59847-C59596. The
XX proteins exhibit neuroprotective, dermatological, immunosuppressive,
XX antiinflammatory, antianaemic, nootropic, antiparkinsonian,
XX cerebroprotective, haemostatic, vulnerary, cytoskeletal, antipsoriatic,
XX antibiotic, virucide, and fungicide activity. The proteins and
XX nucleotide sequences are useful as nutritional sources or supplements
XX and in research. The proteins are useful for treating immune deficiency
XX and disorders, which may be genetic or resulting from infections,
XX autoimmune disorders such as multiple sclerosis, systemic lupus
XX erythematosus, rheumatoid arthritis, and for treating myeloid or lymphoid
XX cell deficiencies such as anaemias by regulating haematopoiesis. The
XX proteins are also useful in compositions for bone, cartilage, tendon,
XX ligament and/or nerve tissue growth or regeneration, for wound healing,
XX tissue repair and replacement and in the treatment of wounds, incisions
XX and ulcers. Other uses include in the treatment of central and
XX peripheral nervous system and neuropathies such as Alzheimer's and
XX Parkinson's diseases and Shy-Drager syndrome, and mechanical and
XX traumatic disorders, such as spinal cord disorders, head trauma and
XX stroke. The proteins may also be used as a contraceptive, and for
XX treating coagulation disorders such as haemophilias. The protein and
XX nucleotide sequences with cadherin activity are useful for treating
XX cancer. Other uses for the protein include for inhibiting the growth,
XX infection or function of, or killing, infectious agents such as bacteria,
XX virus, fungi and other parasites, for effecting bodily characteristics
XX such as height, weight, hair colour, effecting biorhythms or cardiac
XX cycles or rhythms, effecting metabolism, catabolism, anabolism,
XX processing, utilization, storage or elimination of dietary fat, lipid,
XX protein, carbohydrate, vitamins, minerals, cofactors, effecting
XX behavioural characteristics, providing analgesic effects and for treating
XX hyperproliferative disorders such as psoriasis.
XX
XX Sequence 654 BP; 200 A; 183 C; 136 G; 135 T; 0 other;

Query Match 70.4%; Score 630; DB 21; Length 654;
Best Local Similarity 100.0%; Pred. No. 1,4e-309;
Matches 630; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 79 agcatagaccaccagagctgagatctctgacctgagctgcatcccccagagatcaagagccctcc 138
|||||
Db 2 agcatagaccaccagagctgagatctctgacctgagctgcatcccccagagatcaagagccctcc 61
OY 139 agcaggagaacctccatataatctctcaagcaactaacagctgcacacagatctgcgata 198
|||||
Db 62 agcaggagaacctccatataatctctcaagcaactaacagctgcacacagatctgcgata 121
OY 199 aaagtccaatctctccctccctctgtgtgtgcactaaatgtatgtatgtatgtatgtatgt 258
|||||
Db 122 aaagtccaatctctccctccctctgtgtgtgcactaaatgtatgtatgtatgtatgtatgt 181
OY 259 agcagcctgaatccagaggtgtgcacagagccacagagccagagccagagcttcagagga 318
|||||
Db 132 agcagcctgaatccagaggtgtgcacagagccacagagccagagccagagcttcagagga 241
OY 319 tggctccaggaagcgagcccaagaatgtatgataagatctgtctctgagagcccgaga 378
|||||
Db 242 tggctccaggaagcgagcccaagaatgtatgataagatctgtctctgagagcccgaga 301
OY 379 agaaatctcatgacagtgctgt 438
|||||
Db 302 agaaatctcatgacagtgctgt 361
OY 439 ggcattgtgaagaagaagaagacacacacagagccacacagagccacacacagattccaga 498
|||||
Db 362 ggcattgtgaagaagaagaagacacacacagagccacacagagccacacacagattccaga 421
OY 499 gccctgcagcaattctcctaacaatgtcagctcaagaagcttgctgtgtgtgtgtgtgtgtgt 558
|||||
Db 422 gccctgcagcaattctcctaacaatgtcagctcaagaagcttgctgtgtgtgtgtgtgtgtgt 481
OY 559 ctctgagcgccactcttcccaatlaaacattctcagcagaagaagcagtgagcacacta 618
|||||
Db 482 ctctgagcgccactcttcccaatlaaacattctcagcagaagaagcagtgagcacacta 541
OY 619 ccagacactctctctcccaactcactctcccaactgtgacacacccctaaatcattccag 678
|||||
Db 542 ccagacactctctctcccaactcactctcccaactgtgacacacccctaaatcattccag 601
OY 679 tgcctcaaaaagcatgttttcaagatca 708
|||||
Db 602 tgcctcaaaaagcatgttttcaagatca 631

RESULT 11
AAA08343
ID AAA08343 standard; DNA; 849 BP.
XX
XX AAA08343;
XX
XX 30-JUN-2000 (first entry)
XX
XX Human TGC-440 secretory protein nucleotide sequence.
XX
KW TGC-440; secretory protein; immunological disease; infectious disease;
KW pulmonary function disorder; hepatic function disorder; nephrotropic;
KW gastrointestinal function disorder; antiinflammatory; immunomodulatory;
KW virucide; hepatotropic; antianaemic; antibiotic; vaccine;
KW hepatitis; nephritis; influenza; asthma; pulmonary hypertension;
KW pneumonia; Helicobacter pylori infection; ds.
XX
XX Homo sapiens.
XX
XX OS
XX
XX Key Location/Qualifiers
XX 220..579
XX CDS /*tag= a
XX FT /product= "TGC-440"
XX FT /note= "secretory protein"
XX PN WO200014226-A1.
XX

```
PD 16-MAR-2000.
XX
PF 02-SEP-1999; 99WO-JP04765.
XX
PR 03-SEP-1998; 98JP-0250108.
XX
PA (TAKE ) TAKEDA CHEM IND LTD.
XX
PI Itoh Y, Ogi K, Tanaka H, Kitada C;
XX
DR WPI: 2000-256978/22.
XX
P-PSDB: AAY82453.
XX
PT Secretory protein TGC440, antibodies to it and compounds promoting or
PT inhibiting its activity for diagnosis and treatment of diseases of the
XX immune system, lung, kidney, liver and intestinal system
XX
PS Disclosure; Fig 1: 86pp; Japanese.
XX
CC The present sequence encodes a human secretory protein designated
CC TGC-440. TGC-440 has antiinflammatory, nephrotropic, immunomodulatory,
CC virucide, hepatotropic, antilasthmatic and antibacterial activities,
CC and can be used in vaccines. TGC-440 and the polynucleotide sequence
CC encoding it can be used to treat, prevent and diagnose immunological,
CC lung, liver, kidney or gastrointestinal disorders and infectious
CC diseases, such as hepatitis, nephritis, influenza, asthma, pneumonia,
CC pulmonary hypertension, and Helicobacter pylori infection. An antibody
CC immunospecific for TGC-440 is also useful in the above treatment and
CC diagnosis, and also for quantifying the amount of TGC-440 in a liquid
CC specimen.
XX
SQ Sequence 849 BP; 250 A; 219 C; 178 G; 202 T; 0 other;

Query Match 66.3%; Score 593; DB 21; Length 849;
Best Local Similarity 100.0%; Pred. No. 8.3e-291;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaatctgttaacgtaacgaagaacagcgtcacctacactgtgtcttcgcacctaaatggga 60
DB 25 ctatctcgttaacgtaacgaagaacagcgtcacctacactgtgtcttcgcacctaaatggga 84
QY 61 acgctgacctggagtaactaaagcatagaccacagcgtgagtatctcctacttggtatccc 120
DB 85 acgctgacctggagtaactaaagcatagaccacagcgtgagtatctcctacttggtatccc 144
QY 121 caaggatcaggaagcctccagcagggaaacctccatatacttccaagaacttaacgt 180
DB 145 caaggatcaggaagcctccagcagggaaacctccatatacttccaagaacttaacgt 204
QY 181 gaaccagacaggttcgagtgaagaagttcctaactcttcctcctcctcgttgcctcaactatg 240
DB 205 gaaccagacaggttcgagtgaagaagttcctaactcttcctcctcctcgttgcctcaactatg 264
QY 241 ctgagtgcacatgtctctcagcagcctgaatccaaggggtcgcagagggccacagggagcga 300
DB 265 ctgagtgcacatgtctctcagcagcctgaatccaaggggtcgcagagggccacagggagcga 324
QY 301 ggcacagccttcagagataggtctccaaggaagcgccaaagaatgtgagtcaagaattgg 360
DB 325 ggcacagccttcagagataggtctccaaggaagcgccaaagaatgtgagtcaagaattgg 384
QY 361 ttccctgagagcccccagagaagaattcagacagtgtctgggctccaaagaagcagatgc 420
DB 385 ttccctgagagcccccagagaagaattcagacagtgtctgggctccaaagaagcagatgc 444
QY 421 cccctgtgacattcaagggaatgtgaagaacaagacaccaaaggccacacagaag 480
DB 445 cccctgtgacattcaagggaatgtgaagaacaagacaccaaaggccacacagaag 504
QY 481 ccaacaagacattcagagcgtccagcaaatcttccaacaatgtcagctaagaagcttt 540
DB 505 ccaacaagacattcagagcgtccagcaaatcttccaacaatgtcagctaagaagcttt 564
```

```
QY 541 gctctgcttctgagagcgtctgagccaccactctcccaattcaaatctcca 593
DB 565 gctctgcttctgagagcgtctgagccaccactctcccaattcaaatctcca 617

RESULT 12
ID AAF68195 standard; cDNA; 533 BP.
XX
AC AAF68195;
XX
DT 12-APR-2001 (first entry)
XX
DE Human lung tumour protein related nucleotide sequence SEQ ID NO:113.
XX
KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW cytostatic; antisense inhibition; ss.
XX
OS Homo sapiens.
XX
PN WO200100828-A2.
XX
PD 04-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18061.
XX
PR 30-JUN-1999; 99US-0346492.
PR 15-OCT-1999; 99US-0419356.
PR 17-DEC-1999; 99US-0466867.
PR 30-DEC-1999; 99US-0476300.
PR 06-MAR-2000; 2000US-0519642.
PR 22-MAR-2000; 2000US-0533077.
PR 10-APR-2000; 2000US-0546252.
PR 27-APR-2000; 2000US-0560406.
PR 05-JUN-2000; 2000US-0589184.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
PI Reltter MW, Mannion J;
XX
DR WPI: 2001-071488/08.
XX
PT Lung tumor-associated proteins and the nucleic acids that encode them,
XX useful for preventing, diagnosing and treating lung cancer
XX
PS Claim 4; Page 186; 436pp; English.
XX
CC The present invention describes immunogenic portions of lung tumour-
CC associated proteins (I) and the nucleic acids (NAs) that encode them.
CC (I) have cytostatic activity and can be used in gene therapy, antisense
CC inhibition and in vaccines. The NAs and the lung tumour-associated
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with their inappropriate expression,
CC especially lung cancers. For example, the NAs may be administered to
CC treat diseases by rectifying mutations or deletions in a patient's genome
CC that affect the activity of the protein by expressing inactive proteins
CC or to supplement the patient's own production of (I). Additionally, the
CC NAs may be used to produce the lung-tumour associated protein, according
CC to standard recombinant DNA methodology. Conversely, antisense NA
CC molecules may be administered to down regulate protein expression by
CC binding with the cells own genes and preventing their expression. The NA
CC and complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar NA sequences in
CC samples, and hence which patients may be in need of treatment for lung
CC cancer. The (I) may be used as antigens in the production of antibodies
CC and in assays to identify modulators (agonists and antagonists) of the
CC expression and activity of the protein. AAF68083 to AAF68878 and
CC AAF68878 to AAF68878 represent human lung tumour protein related
CC nucleotide and protein sequences which are used in the exemplification
CC of the present invention.
```


KM forensic; gene therapy; chromosome mapping; signal peptide; prostate;
 KM upstream regulatory sequence; cytokine activity; cell proliferation;
 KM differentiation; haematopoiesis regulation; tissue growth regulation;
 KM reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KM thrombolytic; anti-inflammatory; tumour inhibition; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO9906550-A2.
 XX
 PD 11-FEB-1999.
 XX
 PF 31-JUL-1998; 98WO-IB01232.
 XX
 PR 01-AUG-1997; 97US-0905144.
 XX
 PA (GEST) GENSET.
 XX
 PI Duclert A, Dumas Milne Edwards J, Lacroix B:
 DR WPI; 1999-153780/13.
 DR P-PSDB; AAY11732.
 XX
 PT New isolated prostate-derived nucleic acids - used to develop
 PT products which may have cytokine, immune regulatory, haematopoiesis
 PT regulating, anti-inflammatory or tumour inhibition activity
 XX
 PS Claim 1; Page 185; 675pp; English.
 XX
 CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins expressed in prostate, and encode the proteins
 CC given in AAY11716 to AAY11993 respectively. The proteins given represent
 CC the signal peptide and an N-terminal fragment of a secreted protein. The
 CC nucleic acid sequences can be used for producing secreted human gene
 CC products. They can also be used to develop products for diagnosis and
 CC therapy. The proteins obtained may have cytokine activity, cell
 CC proliferation and differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptides can be used for
 CC directing extracellular secretion of a polypeptide or the insertion of a
 CC polypeptide into a membrane, or importing a polypeptide into a cell.
 CC
 XX Sequence 365 BP; 91 A; 100 C; 96 G; 78 T; 0 other;
 SQ
 Query Match 37.4%; Score 335; DB 20; Length 365;
 Best Local Similarity 100.0%; Pred. No. 8e-160;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 368 gagcccgagagaataatcatgacagtgtctgg 402
 Db 331 gagcccgagagaataatcatgacagtgtctgg 365
 RESULT 15
 AAF68125
 ID AAF68125 standard; cDNA; 312 BP.
 XX
 AC AAF68125;
 XX
 DT 12-APR-2001 (first entry)
 XX
 DE Human lung tumour protein related nucleotide sequence SEQ ID NO:43.
 XX
 KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
 KW cytostatic; antisense inhibition; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200100828-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 30-JUN-2000; 2000WO-US18061.
 XX
 PR 30-JUN-1999; 99US-0346492.
 PR 15-OCT-1999; 99US-0419356.
 PR 17-DEC-1999; 99US-0466867.
 PR 30-DEC-1999; 99US-0476300.
 PR 06-MAR-2000; 2000US-0519642.
 PR 22-MAR-2000; 2000US-0533077.
 PR 10-APR-2000; 2000US-0546259.
 PR 27-APR-2000; 2000US-0560406.
 PR 05-JUN-2000; 2000US-0589184.
 XX
 PA (COR-) CORIXA CORP.
 XX
 PI Wang T, Bangur GS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
 PI Retter MW, Mannion J;
 XX
 DR WPI; 2001-071488/08.
 XX
 PT Lung tumor associated proteins and the nucleic acids that encode them,
 PT useful for preventing, diagnosing and treating lung cancer -
 XX
 PS Claim 4; Page 166; 436pp; English.
 XX
 CC The present invention describes immunogenic portions of lung tumour-
 CC associated proteins (I) and the nucleic acids (NAs) that encode them.
 CC (I) have cytostatic activity and can be used in gene therapy, antisense
 CC inhibition and in vaccines. The NAs and the lung tumour-associated
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with their inappropriate expression,
 CC especially lung cancers. For example, the NAs may be administered to
 CC treat diseases by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of the protein by expressing inactive proteins
 CC or to supplement the patient's own production of (I). Additionally, the
 CC NAs may be used to produce the lung-tumour associated protein, according
 CC to standard recombinant DNA methodology. Conversely, antisense NA
 CC molecules may be administered to down regulate protein expression by
 CC binding with the cells own genes and preventing their expression. The NA
 CC and complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar NA sequences in
 CC samples, and hence which patients may be in need of treatment for lung
 CC cancer. The (I) may be used as antigens in the production of antibodies
 CC and in assays to identify modulators (agonists and antagonists) of the
 CC expression and activity of the protein. AAF68083 to AAF68878 and
 CC AAF68648 to AAF68678 represent human lung tumour protein related
 CC nucleotide and protein sequences which are used in the exemplification
 CC of the present invention.

XX
SQ Sequence 312 BP; 100 A; 82 C; 69 G; 61 T; 0 other;

Query Match 34.9%; Score 312; DB 22; Length 312;
Best Local Similarity 100.0%; Pred. No. 3.8e-148;
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 325 caggaaaggcgcccaagaatgtgagtgcaagaatgtgtctctgagagcccgagaagaaaa 384
    |||
Db 1 caggaaaggcgcccaagaatgtgagtgcaagaatgtgtctctgagagcccgagaagaaaa 60
    |||
QY 385 ttcatgacagtgctgtggtgccaagaagcagtgccctgtgatcatlccaaggcgaat 444
    |||
Db 61 ttcatgacagtgctgtggtgccaagaagcagtgccctgtgatcatlccaaggcgaat 120
    |||
QY 445 gtgaagaaaaaacagacccaaggccaagagaagccaagccaagcattccagagcctgc 504
    |||
Db 121 gtgaagaaaaaacagacccaaggccaagagaagccaagccaagcattccagagcctgc 180
    |||
QY 505 cagcaatttctcaacaatgtcagctaaagaagcttgcctgtgcttgtgaagagctctga 564
    |||
Db 181 cagcaatttctcaacaatgtcagctaaagaagcttgcctgtgcttgaagagctctga 240
    |||
QY 565 gcgcccactcttccaaatlaacattctcagccaagaagacaagtgaagacacactaccagac 624
    |||
Db 241 gcgcccactcttccaaatlaacattctcagccaagaagacaagtgaagacacactaccagac 300
    |||
QY 625 actcttctcttc 636
    |||
Db 301 actcttctcttc 312
```

Search completed: September 16, 2002, 06:31:48
Job time: 7656 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 02:53:46 ; Search time 3785.3 Seconds
(without alignments)
1741.436 Million cell updates/sec

Title: US-09-700-770-4
Perfect score: 315
Sequence: 1 taacactgactacgattt.....tgtgtattactactatgc 315

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

| | | | | | | | |
|---|----|------|-------|--------|----|-------------|----------------------|
| C | 1 | 315 | 100.0 | 118068 | 9 | AL356798 | AL356798 Human DNA |
| C | 2 | 40.4 | 12.8 | 226399 | 2 | AC025794 | AC025794 Mus muscu |
| C | 3 | 40 | 12.7 | 125020 | 9 | AF429315 | AF429315 Homo sapi |
| C | 4 | 39.8 | 12.6 | 201794 | 4 | CNS0180W | AL109759 Human chr |
| C | 5 | 39.6 | 12.6 | 62871 | 2 | AC102444 | AC102444 Mus muscu |
| C | 6 | 39.6 | 12.6 | 108873 | 14 | AF318573 | AF318573 Bovine he |
| C | 7 | 39.4 | 12.5 | 54700 | 3 | CEY116A8A | AL117205 Caenorhab |
| C | 8 | 39.4 | 12.5 | 110000 | 2 | CEY113B8.1 | Continuation (2 of |
| C | 9 | 39.4 | 12.5 | 110000 | 2 | CEY113B8.2 | Continuation (3 of |
| C | 10 | 39.4 | 12.5 | 171179 | 2 | AC092859 | AC092859 Pan trogl |
| C | 11 | 39 | 12.4 | 150355 | 9 | HSJ364H10 | AL078603 Human DNA |
| C | 12 | 39 | 12.4 | 168043 | 2 | AC011935 | AC011935 Homo sapi |
| C | 13 | 39 | 12.4 | 174576 | 9 | AC013287 | AC013287 Homo sapi |
| C | 14 | 39 | 12.4 | 185437 | 9 | AC012075 | AC012075 Homo sapi |
| C | 15 | 39 | 12.4 | 218939 | 2 | AC079997 | AC079997 Homo sapi |
| C | 16 | 38.6 | 12.3 | 110000 | 2 | PFMAL13P2_2 | Continuation (3 of |
| C | 17 | 38.4 | 12.2 | 195761 | 2 | AC018895 | AC018895 Homo sapi |
| C | 18 | 38.4 | 12.2 | 203257 | 9 | AC013737 | AC013737 Homo sapi |
| C | 19 | 38.2 | 12.1 | 124281 | 9 | AL136101 | AL136101 Human DNA |
| C | 20 | 38.2 | 12.1 | 173452 | 9 | AC003013 | AC003013 Human PAC |
| C | 21 | 38 | 12.1 | 9577 | 1 | U67580 | U67580 Methanococ |
| C | 22 | 38 | 12.1 | 68464 | 9 | AC005248 | AC005248 Homo sapi |
| C | 23 | 38 | 12.1 | 169469 | 2 | AC037195 | AC037195 Homo sapi |
| C | 24 | 38 | 12.1 | 172144 | 4 | AC092497 | AC092497 Sus scro |
| C | 25 | 38 | 12.1 | 175946 | 9 | AP002088 | AP002088 Homo sapi |
| C | 26 | 38 | 12.1 | 181686 | 2 | AC097819 | AC097819 Rattus no |
| C | 27 | 37.8 | 12.0 | 120984 | 9 | AC006463 | AC006463 Homo sapi |
| C | 28 | 37.8 | 12.0 | 175735 | 2 | AC068566 | AC068566 Homo sapi |
| C | 29 | 37.8 | 12.0 | 192120 | 2 | AC104640 | AC104640 Homo sapi |
| C | 30 | 37.6 | 11.9 | 14059 | 3 | AE001390 | AE001390 Plasmodiu |
| C | 31 | 37.6 | 11.9 | 36255 | 2 | AC016208 | AC016208 Homo sapi |
| C | 32 | 37.6 | 11.9 | 88037 | 2 | PFMAL13P8 | AL096782 Plasmodiu |
| C | 33 | 37.6 | 11.9 | 146389 | 2 | AC023916 | AC023916 Homo sapi |
| C | 34 | 37.6 | 11.9 | 166703 | 9 | AL139378 | AL139378 Human DNA |
| C | 35 | 37.6 | 11.9 | 167257 | 9 | AC008808 | AC008808 Homo sapi |
| C | 36 | 37.6 | 11.9 | 167553 | 9 | AC090042 | AC090042 Homo sapi |
| C | 37 | 37.6 | 11.9 | 169480 | 2 | AC068074 | AC068074 Homo sapi |
| C | 38 | 37.6 | 11.9 | 182651 | 2 | AC104580 | AC104580 Homo sapi |
| C | 39 | 37.4 | 11.9 | 104089 | 9 | AC012488 | AC012488 Homo sapi |
| C | 40 | 37.4 | 11.9 | 112212 | 9 | AL136096 | AL136096 Human DNA |
| C | 41 | 37.4 | 11.9 | 139226 | 2 | AC084233 | AC084233 Homo sapi |
| C | 42 | 37.4 | 11.9 | 166946 | 2 | AP001008 | AP001008 Homo sapi |
| C | 43 | 37.4 | 11.9 | 222029 | 2 | AL606831 | AL606831 Mus muscu |
| C | 44 | 37.2 | 11.8 | 12240 | 3 | AE001424 | AE001424 Plasmodiu |
| C | 45 | 37.2 | 11.8 | 167491 | 2 | AC093442 | AC093442 Papilio cyn |

ALIGNMENTS

RESULT 1
LOCUS AL356798/c
DEFINITION Human DNA sequence from clone Rpl1-467B11 on chromosome 9, complete sequence.
ACCESSION AL356798
VERSION AL356798.18 GI:15795413
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 118068)
AUTHORS Babbage,A.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
REQUESTS: clone requests: sanger.ac.uk
COMMENT On Sep 27, 2001 this sequence version replaced gi:14455883.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/GRP/Ch9> RP11-467B11 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/Bacpac/home.htm> VECTOR: PBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-467B11 it may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone RP11-467B11 is at 118068 in this sequence. The true left end of clone RP11-208F1 is at 29578 in this sequence. The true right end of clone RP11-29B11 is at 2000 in this sequence.

FEATURES
source
1. 118068
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-467B11"
/clone.lib="RP11-11.2"
/clone.lib="RP11-11.2"

BASE COUNT 33053 a 23625 c 22510 g 38880 t
ORIGIN

Query Match 100.0%; Score 315; DB 9; Length 118068;
Best Local Similarity 100.0%; Pred. No. 1.3e-62;
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taacacgactagaatttaagaataaacttttgagaatagaaacaaatgaatgaatt 60
Db 96845 TAACACTGACTGATGATTTAAAGAAATTAACCTTTGAGAAATAGAACAAATGAAATCAGTT 96786
QY 61 tctccaccacttaagatatactctcttagagatctaacagcctcccttagagggaatacaaa 120
Db 96785 TCTCCACCACTTAAGATATATCTTTAGAGATCTACAGCCTCCCTTTAGGGGACATACAAA 96726
QY 121 gtcagttgtgtgcttcttlttgagtcacacactatataatgaagtagatgaactacaatt 180
Db 96725 GTCAGTTGTTGCTTCTTTGTTGAGTCCACCTTATTTCAAGTAGTATGACACCAATT 96666
QY 181 ttgaataatgattgtacacaaataactgtgagttatggaaacatagtagaagaataa 240
Db 96665 TTGAATAATGATTGTACACAAATAAAGTATGAGAAACATCAGTAGAAGAAATA 96606
QY 241 caacattcaccccttaagagatcattacttgaactagagatatttctctgtgt 300
Db 96605 CAACATTCATCCCTTTACAGAGATCATTTACTTGCAACTCAGGATTAATTTGTCATGTGT 96546
QY 301 attatctactatgac 315
Db 96545 ATTATCTACTATGCG 96531

RESULT 2
AC025794 226399 bp DNA linear HTG 13-JUN-2001
LOCUS AC025794
DEFINITION Mus musculus clone RP23-6207, WORKING DNAfr SEQUENCE, 25 unordered

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
1 (bases 1 to 226399)
Mus musculus, clone RP23-6207
2 (bases 1 to 226399)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bida, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Haagos, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melgrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaferr, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zalnoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 13, 2001 this sequence version replaced gi:8225449.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information

Center project name: L8057
Center clone name: 62_0_7
Summary Statistics
Sequencing vector: M13; M77815; 45% of reads
Sequencing vector: Plasmid; n/a; 55% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 217695 bases at least Q40
Consensus quality: 221726 bases at least Q30
Consensus quality: 223237 bases at least Q20
Insert size: 219000; agarose-fp
Insert size: 223999; sum-of-contigs
Quality coverage: 8.2 in Q20 bases; agarose-fp
Quality coverage: 8.0 in Q20 bases.
NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 15086: contig of 15086 bp in length
* 15087 15186: gap of 100 bp
* 15187 15923: contig of 737 bp in length
* 15924 16023: gap of 100 bp

```
* 16024 16694: contig of 671 bp in length
* 16695 16794: gap of 100 bp
* 16795 17880: contig of 1086 bp in length
* 17881 17980: gap of 100 bp
* 17981 18573: contig of 589 bp in length
* 18580 18679: gap of 100 bp
* 18680 19268: contig of 589 bp in length
* 19269 19368: gap of 100 bp
* 19369 20014: contig of 646 bp in length
* 20015 20114: gap of 100 bp
* 20115 20821: contig of 707 bp in length
* 20822 20921: gap of 100 bp
* 20922 21533: contig of 612 bp in length
* 21534 21633: gap of 100 bp
* 21634 22538: contig of 905 bp in length
* 22539 22638: gap of 100 bp
* 22639 22338: contig of 700 bp in length
* 23339 23438: gap of 100 bp
* 23439 24125: contig of 687 bp in length
* 24126 24225: gap of 100 bp
* 24226 24941: contig of 716 bp in length
* 24942 25041: gap of 100 bp
* 25042 25887: contig of 846 bp in length
* 25888 25987: gap of 100 bp
* 25988 26876: contig of 889 bp in length
* 26877 26976: gap of 100 bp
* 26977 28437: contig of 1461 bp in length
* 28438 28537: gap of 100 bp
* 28538 30287: contig of 1750 bp in length
* 30288 30387: gap of 100 bp
* 30388 32661: contig of 2274 bp in length
* 32662 32761: gap of 100 bp
* 32762 36782: contig of 4021 bp in length
* 36783 36882: gap of 100 bp
* 36883 45293: contig of 8411 bp in length
* 45294 45393: gap of 100 bp
* 45394 61819: contig of 16426 bp in length
* 61820 61919: gap of 100 bp
* 61920 126287: contig of 64368 bp in length
* 126288 126387: gap of 100 bp
* 126388 149055: contig of 22668 bp in length
* 149056 149155: gap of 100 bp
* 149156 211213: contig of 62058 bp in length
* 211214 211313: gap of 100 bp
* 211314 226399: contig of 15086 bp in length.
FEATURES
    source
        1..226399
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone_lib="RP23-6207"
            /clone_lib="RPC1-23 Female Mouse BAC"
            1..15086
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left"
misc_feature
    15187..15923
        /note="assembly_fragment"
misc_feature
    16024..16694
        /note="assembly_fragment"
misc_feature
    16795..17880
        /note="assembly_fragment"
misc_feature
    17981..18579
        /note="assembly_fragment"
misc_feature
    18680..19268
        /note="assembly_fragment"
misc_feature
    19369..20014
        /note="assembly_fragment"
misc_feature
    20115..20821
        /note="assembly_fragment"
misc_feature
    20922..21533
        /note="assembly_fragment"
misc_feature
    21634..22538
        /note="assembly_fragment"
```

```
misc_feature 22639..23338
              /note="assembly_fragment"
misc_feature 23439..24125
              /note="assembly_fragment"
misc_feature 24226..24941
              /note="assembly_fragment"
misc_feature 25042..25887
              /note="assembly_fragment"
misc_feature 25988..26876
              /note="assembly_fragment"
misc_feature 26977..28437
              /note="assembly_fragment"
misc_feature 28538..30287
              /note="assembly_fragment"
misc_feature 30388..32661
              /note="assembly_fragment"
misc_feature 32762..36782
              /note="assembly_fragment"
misc_feature 36883..45293
              /note="assembly_fragment"
misc_feature 45394..61819
              /note="assembly_fragment"
misc_feature 61920..126287
              /note="assembly_fragment"
misc_feature 126388..149055
              /note="assembly_fragment"
misc_feature 149156..211213
              /note="assembly_fragment"
misc_feature 211314..226399
              /note="assembly_fragment"
              clone_end:T7
              vector_side:right"
```

BASE COUNT 61164 a 51232 c 50642 g 60960 t 2401 others

ORIGIN

Query Match 12.8%; Score 40.4; DB 2; Length 226399;
Best Local Similarity 54.8%; Pred. No. 5.6;
Matches 80; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
QY 10 actcagatttaagaacttcttgaagaataagaacaatgaatcagttccacaca 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51932 ATTCGCCATTTCGGAAATGCAATTCACACAGTCTTCAGATGATGCTTCTTCACCA 51991

QY 70 cttaagtatactcttaagatctacagctcccttccttgaggactacaagtcagttgt 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51992 CTGAACAATACCTCAAAATATTCATTACCCACACCACTTGATGACCAATTCAGTGGT 52051

QY 130 gtgccttggtgagtcacacctat 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52052 CTTGGCCTTGAAAAAGCCAGTGTCA 52077
```

RESULT 3
AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315.1 GI:17646244
VERSION AF429315.1
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 125020)
Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll,Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
JOURNAL MEDLINE 21583737
PUBMED 11694876

```

REFERENCE      2 (bases 1 to 125020)
AUTHORS        Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
TITLE          Direct Submission
JOURNAL        Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
                Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES
SOURCE
1.125020
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="Isolated from a patient with Huntington's
disease-like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit=cig
complement(<36507..>36887)
/gene="JPH3"
/product="junctionphillin 3"
complement(<36507..>36887)
/gene="JPH3"
/note="JPH3"
complement(<36507..>36887)
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctionphillin 3"
/protein_id="AA140941.1"
/db_xref="GI:17646245"
/translation="MSSGGRFNPDGSGYCGGMDGKAGHGVCTGPKGRTGTSWS
HGFEVLGVYTPSGNTGYOSTMAQGRKHGIGLSKGRKWTGKGTGFGRTGVRCAQ
NGAKYEGTWSNGLDQGTETSTSDG"

BASE COUNT    29056 a 32731 c 30696 g 28283 t 4254 others
ORIGIN
Query Match      12.7%; Score 40; DB 9; Length 125020;
Best Local Similarity 13.7%; Pred. No. 7.6;
Matches 34; Conservative 112; Mismatches 103; Indels 0; Gaps 0;

QY 27 taactttgagaataagaacaaatgaatcagtttccaccacctaagatactctta 86
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66237 KMTTKTTTGTGTRGRRSARSSMYVMSCTWYSTTSKYVSSCTGKTYTTKAWTKCCTKMM 66296
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 87 gagatctaacagcctccttagggagacatacaagtcagttgtgtccttggtagtc 146
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66297 MKCMSYSATYCWYTYTTRKYYTYTTRAAWSMCSYSRSMKWRMSRCMYRMTSTMSG 66356
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 ccacctatattcaagtagtgtagtactacaatttgaataatagattgcacacataaa 206
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66357 NAKSTRRCRGWTWMSMTMKMKMKMKMKMGWYTRTMAAAMWYKTRRMAAAMWMA 66416
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 ctggaagttatagaacacagtagaagaatacaacatccatcccttacagagatc 266
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66417 TTKYMKKKRGMGARCMAAMAMWYARMKMKWCTKTCATMYSYWMMKRTGTYRMC 66476
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 267 attacttg 275
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66477 YTYCWECTR 66485
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
CNS0180W
LOCUS          CNS0180W      201794 bp      DNA      linear      PRI 04-MAY-2001
DEFINITION    Human chromosome 14 DNA sequence BAC R-898B23 of library RPCI-11
                from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION     AL109759
VERSION       AL109759.4      GI:8176900
KEYWORDS      HTG.
SOURCE        human.
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE      1 (bases 1 to 201794)
AUTHORS        Helling,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
                Brothier,P., Catolico,L., Barbe,V., Pelletier,E., Ariguenave,F.,
                Levy,M., Eckenberg,R., Bruls,T., debernardinis,V., Cruaud,C.,
                Gypay,G., Saurin,W. and Weissenbach,J.
TITLE          Sequencing of the human chromosome 14
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 201794)
AUTHORS        Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (04-MAY-2001) Genoscope - Centre National de Sequencage :
                BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
                On Jun 2, 2000 this sequence version replaced gi:6434630.
COMMENT        ----- Genome Center
                Center: Genoscope / Centre National de Sequencage
                Center code: GS
                Web site: http://www.genoscope.cns.fr/
                Contact: SeqRef@genoscope.cns.fr
                -----
                The following BAC sequence is oriented from the T7 to the SP6 end.
                Upstream BAC (overlapping the T7 end) : R-35609 (AC-AL121790)
                Downstream BAC (overlapping the SP6 end) : R-158A24
                Summary Statistics
                Assembly program: Phrap; version 1.0
                Quality coverage: 6.74x in Q20 bases; sum-of-contigs
                -----
Overall quality chart :
Range      : bases
0 - 9      : 24
10 - 19    : 210
20 - 29    : 1378
30 - 39    : 3593
40 - 49    : 11245
50 - 59    : 36298
60 - 69    : 72785
70 - 79    : 50952
80 - 89    : 25309
90 - 99    :

Percentage of bases with a quality value >= 40 : 99 %.
FEATURES
source
1.201794
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_id="R-898B23"
/clone_lib="RPCI-11"
61396..61519
/note="matching EMBL:AA128898
RHdb:RH48004
dbSTS:STS41063
Identified using the e-PCR software (G. Schuler)"
201134..201288
/note="matching EMBL:AA012815
RHdb:RH75590
dbSTS:STS52678
Identified using the e-PCR software (G. Schuler)"

BASE COUNT    69076 a 36481 c 35544 g 60693 t
ORIGIN
Query Match      12.6%; Score 39.8; DB 9; Length 201794;
Best Local Similarity 52.8%; Pred. No. 7.8;
Matches 86; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 33 ttgagaataagaacaaatgaatcagtttccaccacttaagatatactcttagatc 92
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67232 TAGATAATAATAGAAATATTTCATCACCCCTACATATTCCTTACTTACTATC 67291
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Oy 93 tacagcctcccttaggggacatacaagtcagtgctgtgcttctgttgagccacct 152
 Db 67292 TTCACCTTCACTAAGGATGATCATCTTGACTCTGATCATAGATCATCTTGCT 67351
 Oy 153 tatatacagtagtgatgactacaatttgaataatgattgt 195
 Db 67352 GTTATTGAACCTTATCATCATATATATATATATCTCT 67394
 RESULT 5
 AC102444/c 62871 bp DNA linear HTG 23-NOV-2001
 LOCUS Mus musculus clone RP24-196115, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION AC102444
 AC102444.1 GI:17061530
 VERSION HTG: HTGS_PHASE0.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 62871)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 JOURNAL Mus musculus, clone RP24-196115
 REFERENCE 2 (bases 1 to 62871)
 AUTHORS Unpublished
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Bouckgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Chepel, I., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Haas, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kanat, A., Karatas, A., Kells, C., Laroque, K.,
 Lamazares, R., Landers, T., Lenockzy, U., Levine, R., Liu, G.,
 Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
 Meneses, T., Miho, V., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigliio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M. J., Young, G.,
 Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L18811
 Center clone name: 196_1_15

 * NOTE: This record contains 84 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 626: contig of 626 bp in length
 627 726: gap of 100 bp
 727 1387: contig of 661 bp in length
 1388 1487: gap of 100 bp
 1488 2127: contig of 640 bp in length
 2128 2227: gap of 100 bp
 2228 2865: contig of 638 bp in length
 2866 2965: gap of 100 bp
 2966 3591: contig of 626 bp in length
 3592 3691: gap of 100 bp
 3692 4331: contig of 640 bp in length
 4332 4431: gap of 100 bp
 4432 5087: contig of 656 bp in length
 5088 5187: gap of 100 bp
 5188 5836: contig of 649 bp in length
 5837 5936: gap of 100 bp
 5937 6580: contig of 644 bp in length
 6581 6680: gap of 100 bp
 6681 7335: contig of 655 bp in length
 7336 7435: gap of 100 bp
 7436 8093: contig of 658 bp in length
 8094 8193: gap of 100 bp
 8194 8836: contig of 643 bp in length
 8837 8936: gap of 100 bp
 8937 9589: contig of 653 bp in length
 9590 9689: gap of 100 bp
 9690 10312: contig of 623 bp in length
 10313 10412: gap of 100 bp
 10413 11061: contig of 649 bp in length
 11062 11161: gap of 100 bp
 11162 18183: contig of 652 bp in length
 11814 11913: gap of 100 bp
 11914 12561: contig of 648 bp in length
 12562 12661: gap of 100 bp
 12662 13507: contig of 646 bp in length
 13508 13407: gap of 100 bp
 13408 14046: contig of 639 bp in length
 14047 14146: gap of 100 bp
 14147 14802: contig of 656 bp in length
 14803 14902: gap of 100 bp
 14903 15545: contig of 643 bp in length
 15546 15645: gap of 100 bp
 15646 16288: contig of 643 bp in length
 16289 16388: gap of 100 bp
 16389 17032: contig of 644 bp in length
 17033 17132: gap of 100 bp
 17133 17798: contig of 666 bp in length
 17799 17898: gap of 100 bp
 17899 18526: contig of 628 bp in length
 18527 18626: gap of 100 bp
 18627 19296: contig of 670 bp in length
 19297 19396: gap of 100 bp
 19397 20033: contig of 637 bp in length
 20034 20133: gap of 100 bp
 20134 20764: contig of 631 bp in length
 20765 20864: gap of 100 bp
 20865 21496: contig of 632 bp in length
 21497 21596: gap of 100 bp
 21597 22252: contig of 656 bp in length
 22253 22352: gap of 100 bp
 22353 22994: contig of 642 bp in length
 22995 23094: gap of 100 bp
 23095 23763: contig of 669 bp in length
 23764 23863: gap of 100 bp
 23864 24520: contig of 657 bp in length
 24521 24620: gap of 100 bp
 24621 25282: contig of 662 bp in length
 25283 25382: gap of 100 bp
 25383 26040: contig of 658 bp in length
 26041 26140: gap of 100 bp
 26141 26795: contig of 655 bp in length
 26796 26895: gap of 100 bp
 26896 27535: contig of 640 bp in length

[illegible]

| | | | | |
|-----------------------|-----------------|-----------------|-----------|---------------|
| Query Match | 12.6%; | Score 39.6; | DB 2; | Length 62871; |
| Best Local Similarity | 49.5%; | Pred. No. 11; | | |
| Matches 102; | Conservative 0; | Mismatches 104; | Indels 0; | Gaps 0; |

Qy 1 taacactgactcagattttaagaatacaacttttgagaatagacaacatgcaatcagtt 60
 ||| ||| | | | | | | | | | | | | | | | |
Dd 32693 TAACCAAGTATATAATTTCACAATAATTAACTAAAAATTAAATTAAATAAAAAT 32693

Oy 61 tccacacactaagatatactctcttagagatctaaagcctccctttaggggaacatacaa 120
 Db 3263 TTGAAATTATTAATAATTCTTAATATGTGTTAACTTCTGCCCTACTCTTAATTAACAT 32574
 Oy 121 gtacgttgcttgcccttgtagtccacactatattcaagtagtgaactacaatt 180
 Db 32573 TATATGACATGTTTTATGATTAAATCAACCACTTAATTAATAAAATGTTCATATATCAT 32514
 Oy 181 ttgaanaatagattgtccacaataa 206
 Db 32513 TTCCGAATATTGTTCTATCAACAACAAA 32488

| RESULT | 6 | | | | | |
|------------|-----------------------|---------------------|-----|--------|--------------------|--|
| AF318573 | | | | | | |
| LOCUS | AF318573 | 108873 bp | DNA | linear | VRL. 18-MAY-2001 | |
| DEFINITION | Bovine herpesvirus 4 | long unique region, | | | complete sequence. | |
| ACCESSION | AF318573 | AF271211 | | | | |
| VERSION | AF318573.1 | GI:12802528 | | | | |
| KEYWORDS | | | | | | |
| SOURCE | Bovine herpesvirus 4. | | | | | |

REFERENCE
AUTHORS
1 (bases 1 to 108873)
Zimmermann, W., Broll, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and

TITLE Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and Identification of an Origin of DNA Replication
JOURNAL J. Virol. 75 (3), 1186-1194 (2001)
ISSN 0022-5380

| | |
|----------|---|
| AUTHORS | Zimmermann, W., Broll, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and Goltz, M. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (03-NOV-2000) Project P24, Robert Koch-Institut, Nordufer 20, Berlin D-10555, Germany |
| FEATURES | Location/Qualifiers |
| source | I..108873 |

```
misc_feature 1. .108873 /note="long unique region; LUR"  
CDS          complement(445. .633)
```

```

/evidence=not_experimental
/product="hypothetical protein pbo1"
/protein_id="AAK07921.1"
/db_xref="GI:12805259"
/translation="MGGVSNVAFGRRTQDILNSAGNEVCAVLEREKHNHNSAGC
CAVRRNNSPFHSLPGMP"
614..4480
DSDS

```

CDS

```

/note="segment protein; ORF3 BOREAL"
/codon_start=1
/product="γ-FAM-synthase"
/protein_id="AAK07922.1"
/db_xref="GI:12802530"
/translation="MDPTSTQLCFEVSRRAPLNSNCCQVLIQPTSGQLPTSPVGLP  

EDEVVAVLGEQGAARFLVPLPLKTSLSGELTPRESSILTDKRSGLFVSPGSA  

PTALSRLVLIILSMGLSALNPLNLTQILVSLITLCKNLNPSNLSHSTRAL  

CPSPSLFPRYGSYISGAFPTGAPGAPGQVLSPEKDLGALVGGTILNLSGASGEFT  

POHGQITVQNECKTAYIVDVIPILEISTYLSGHMTDGLAVYVGGIQLQVAFRN  

LKLKLYSNPQNEFNVSEYSSQLSIASTVYITSPSESLPEPTYDMLDAVIMRCC  

SLPQCGFFVYFVSHSLPLVLAQNSPTFAGTGAACSCRNITDLPETNRKINILQCT  

FTSSHESGSEFRCVLQALVLAQNSPTFAGTGAACSCRNITDLPETNRKINILQCT  

IYQSLAPLPLMTGLLPTLASLNQNSPTFAGTGAACSCRNITDLPETNRKINILQCT  

AGCAHGGQGVYGRVGRPHITHTYKDEGQVREKISIKRSTSEPLSTLNKIQOS  

ENMQDSNMQDTPSLSYTLLLEILKIPANGCCQITVXKHIDRLSSRGAQOQIGCPIS  

PISDSLINSDLSLAAPVQDSNPNWPAEESLDLPLVNNNEFVPGICSAAGCEPLS  

YTFPKGHAVALVSTSLNIVSAFKRIIDVQTEVNIWPPQDTSFGILELMRAGRE  

TSQALGIVFVSTSLNRRKQQLINNSLVRLVLAIVAPCKDVQDGLPDILKEDPDS"

```

| | | |
|-----------------------|---|--|
| CDS | WLPSTSEYHTEGTVISQLFRDSPSGHIIINIDPLVYKKLITVTKHLINHSEIISCHDV GGGLITACFEKMAVAGASIALTVPODEDVPLFSETPGALVEVPRKVCSTYOKHLE TSDIIFEDVGRPLPSVANTPELVSYKEBRLIRRESINEMVENNRHSSTYKQRTYCEY OGTPEDPKDQLHLITFOPSMCPHGPYTHOVNVLPLPGTINPDSLVALEBAGRVLN LVSTETDVKVITVDTTVNFSGILGANIDBATGDKAISMYTKHNSVLGELKLI NSPVSFLAIGHACQILFENKFMGNKSPNTYCKENYSGIMESRNINFEIPENHI AVAQSMKSGSLPGWIOGTHLGFHAFSELYEMMLSTHGMVAIOFGADISAPALTYP ONPRTYGISGLCSADGRHLALHDPGLSNNLMQMPHILPKMPPILAKVSPKRMFLDLH IWANKVREMDOPRPHPDPLNRINIKYM" | 4608..4811 |
| CDS | /note="ORF B02; B0RFA2" /codon_start=1 /evidence="not_experimental" /product="hypothetical protein pb02" /protein_id="AAK07923.1" /db_xref="GI:12802531" | |
| CDS | /translation="MASPTQYCHRMVRYHGRKSAKISYGVKDRDYTHQLSKNSAYS -GMSDQPCSCADLNKLFVSQCPS" 5286..5513 /note="ORF B03" /codon_start=1 /evidence="not_experimental" /product="hypothetical protein pb03" /protein_id="AAK07924.1" /db_xref="GI:12802532" | |
| CDS | /translation="MGFNPVMLYFHCNCTPISLKHGLYIKIKIISYLIRATFLVRN HIGPKRDPTRPGCGRPPSLIVGRALKK" 5943..9347 /note="ORF 6" /codon_start=1 /product="single-stranded DNA-binding protein MDBP" /protein_id="AAK07925.1" /db_xref="GI:12802533" | |
| CDS | /translation="MAAKTAQCGQYEDNIGSKAPRGPGCIYIVYKKEFPFKASLIG NGEPGTSVSWPPLFGLVPEGPIINVAVHKIDTTVYSAKVITSYHREVMESNIDC ETPIERHGLPOLCOETRDLEFGFIHSPSPSNTNNISLQDLGTCNCDIEWLIGVIT EGFKEPLHYGHMIPILGHVEEVOISPOHAAKRIPLYDEDLFSKIAPDMKRFHSYVS OYLENSLYTAIAQAIRVKDVAIVIOAMSOFRQDHMKPKVYHKRCPSSSRGPDGL SLMIDVSAELAVSYGLSFIDAPQDMSALIDYTSMPYFVDCITEEDRLQALHAANK OSIHYNPOLFTSNMLYTRIDOKONPKAKGDLANNYSYILOHGISYLSBAQDQYCG PVFGVQSNLSGSTYTTIHLNLAIVASMCENLARICTYLQFCOHKSTONOSYINSDY VGTAAASDMCNLCQGRKPAVCINTLFYRLRDFPIITISHRDPIITGSVGMNIDLD ILGNFASFREKEDEGAQVEVOKYTMQTLLEKLEGMGKIDTPPEKLSLIDPSE IKVEFDIDAHVDEVLKFINCKVKNINFERETIKSIHHYIOYCNVFAOPPCPVLOL YKSKLTIIDOCYLCNMAWYEDDSSMGMPTEMLKMHYOTLMTNKKCCPEPRGLTG AEKYMHSDDPCDFPDVDAINGTISPTKTQVIRISRAMVYKTIKIKRRIIFSSNSG SEAIOSGLFSGSKKDNITVTEGPIKFLNITHKIMFPQTKSSLFMMHFTSTKQIPL ITGVPRDQLVALANIIEYNSKLHSEIDVLDIIPDNLITYAKIKLNNAILRACGQYQY STTHCLVPTVATVGEERYPHVLENEEFGTBOYLKLVONRTAQIVQATLKEDVAMQ KLRLIITVPMVINKYTGINGNNGIFHCNGLGYFMGRVDRNLIFENAPRPOSTAVM RKHYEMTPIYDNLIKRTVPPSSSTFEVENTIRDRMLLEDKDNONITFRDYLDELVG LGOACADITADIDLOFYIGEYVIMSDILSRQSIDACVPMCEDSVSSSLGELYTSEE OLEFTGLSEDTTSAPTSEDFQASGLSTIAAGKRRKLKMLCMISDFL" | 9352..11430 /note="ORF 7" /codon_start=1 /product="transport protein" /protein_id="AAK07926.1" /db_xref="GI:12802534" |
| CDS | /translation="MARELAAYSQVDFLSFKTSLITFCDPRHIDLGAFQRRKRLKL LIDALYSHLVEONRTESCISLELHLIINTYITMQKLETENINIEYDPAAMLHS DLCEFHKNVNLCEYGDVYIPISLIINDIEIFRSNSVFCYSINNALCEGLEETINF LGTLRNISPRLPOLYIPISCMCNLETSILPONGENVFTIMNSINCHVATPYAPE PYQGLFENELMHLNITPVGTDKVOLETSDRTIQNASIEAIDOHNIIEKVSCHITLIS NLITYNSGNSDQPSRLSCQMAWLIQHTFRKASRQVASHALATTPRIHFEDSYRA YPIESLFGGIIIFYIMDNIQALKDCSTFFLAKSNYKTIEMERKNELVYANLLASH TKESGIDKPNISIKOARHTIIGETSEOFSAARKNNYLOKYTEGKKTIDTLET OGKIMCNTLLRTGWMVYKKAAILKNHFMIRKRIALDPMHNTYATINELYESKFI KNSLYVOKISOEHIDISTIQYSLITGPIVESPDPFPPSPNILLANCDAAKVMFHKK MLISEVIMPSMEKDMIDAEEFNAPYINIGKTLNSVQOCAMRYIRELVMSVSLYNVWE KDLNIVSPSTLSPHMLFSKPTFKSGIYIYKTAPLLFIHNTQMIFKDLVALLYHH IOLTSHKNVL" | 11420..14044 |
| gene | | |
| CDS | /gene="gb" 11420..14044 /gene="gb" /note="ORF 8" /codon_start=1 /product="glycoprotein B" /protein_id="AAK07927.1" /db_xref="GI:12802535" | |
| CDS | /translation="MYKTLIFLAIKVCSPQNTTHSTTSPTSISSTTSSTTSK PSNTYSTNSLSLASSPNTSTSPSTDNQGTSPITPTVYTDASNKTRKVCSSSS SGELFRFDLDQCPDPDKKHVEGILLVLAKNIIVYIERKVRKRIATSVYRWGMS AAVTNRDDISRAIPYNEISMIDRIYHCSAMATVINGLINTIDDSBNKSPLOQPA GLTENINREISOPLIYABEGPGLIYRRTVYNCVVMVAASVPYRHHFIALDPTI EISPCRHNSOCTGTGNSRQATKVMIENHOTVDYERGGPTKDKRIPIKDEEYTS WKAEDREAIQDFVYKTPPRALQIOTIHNSHFVANEVYASLTISNOBETLRGTEI LNCMNSTINETLEETVKKFMKSHIDGEEKYIKTNGSLFLMQAKPPLASBHTYTI ERNNTGNKRSROKRSVDKRTQGAAGLSYAOQVAYDHLRTSMNHIIEBLRTWCRQ KKNLNMWELSKINPVSMAALYGGPVAVKAGDAFMYSECINVOASVINIKSMRTD DPKVCYSRPLVTEFEVNSTAFPRGOLGRNELINTNTVENCVPADAHYFVKNMTHY FKDYKFEVMTQNNISITLDTFLTLNLFIDNIDFTEVELYESTERKMSALDLETFR EYNYTOKLASREDMDNTIDNRDLKQDISEMADGQIDCKGVYVNTFSGIVYVGS IVGGVSEFTNPIGCVTITLILLYVVEVYFYSRPTNNKNEAPIKMTIPNIDKASEQ NIQPLGGEIKRILGLMHQLOOSEHGKSEESKSPGLFOLLGDELQILRRGYRLP TFDPSPGNDTSETHQYV" | 14184..17201 /gene="Dpol" 14184..17201 /gene="Dpol" /note="ORF 9" /codon_start=1 /product="DNA polymerase" /protein_id="AAK07928.1" /db_xref="GI:12802536" |
| CDS | /translation="MEFNPTYSKKTTRDKRISCVDDPDKTINPYRLVPCKPTGCP GVISITTTDTFVLFHQDKVEPLTSGDQORSWSLAKPPAIPKDKLFEVYDIIIE TITYDREPDPIQOTDIPNGTVALIKGLQODASVCVNFQOMVYFVYLVPGVN LSVYLOQTLGVNGVKNQCTKESITQORRKILKRYDLSLIPVYITLSAPTEINQVLN TSCGCEVESVNSASTRFIDNKFSTFGVYSNPNRFTQSDSTPDSFQDLQ FHEEQQMPPRYIMSFIDIECIGEGPFCATKREDIVIOISCIIMKVCSSPPNILLS LGTCPRIANTDYEPSELDMFATFTLLRDNIDPVGYNINANDPPIITDRAOY NIPRHFRTKLVISGSTFEVQNPMTGAGFWRSKIRISQVLDIMQYCKDLMSDY KLNTYAKHQMGOEDVYKKEIPSLFRSGEASRARIGSCVLDVSLVLDLKYFMIHV EISEIAKMAKIPARVLLNDGQOIRVFSCLLEAKKEHFLIPGAKRPGQGVATVIN PRPGYVPIIIVDPAASYPSIIQAHNSCLYSPMIDONLHLHLKPDYEFTHLSTGP IHFVQKHOTKLSITLITLMTAAMKARKRTIKELANCDDGPMKTTILDKQOLAIKYTCMSVY GFTYASGLPCITPIAEVYTLQGRIMLEKSKAFVEMITPERLSIDVSYVPDPPASP RVYIGDTDLFTECGYEPQOSVLANADKLTEYTKALFEDPIKLAERTYQCLMLK KRYIGILSDSKLVMGVLDLRNTACSPVQTSKEILDVLRDPEVQAQAYCGRQAPA KYVSGLPEVGEKVIDVNLNSYAAALATGVSESLTFSTELSPRGEYKTTLLPHLTV YRKINSRNEELPOLHRIIPYVFIKDDORCKEDLADDPYVQNKIPISEVIEYFDKLI HGANVILQCLFQENSGNMVYEILYNVNIYPSFT" | 17297..18577 /note="ORF 10; B0RFB1; conserved in other gamma-herpesviruses" /codon_start=1 /product="unknown" /protein_id="AAK07929.1" /db_xref="GI:12802537" |
| CDS | /translation="MTSLVERFSISGMVONICNGIILMTNLKDIPOVGGTGLYKVV FLPLSLQNNIYISRGVQSIIELEDAQACRAMVQCFEPIRPLTEPIIHLVEN MPLFLVQNTIINSLSLAIIPMKPTCPGTCISYSTAPETIIPQSHANHLPS TDQGHLLTSGQAIKCNHKRYTVPFRGSPMOEFGMFKWASRGRLQDACVGMVSYD IKNIMAEVYNPAGIKKYSVINTPHPRQHIWTLKFTYKVMKTRDILVASYLPSLPS CGKVMNIITDSERKILKPGSTLMLKLYITTRGNESTKAMVFTTGTNTNPMVITTEPI LPMPTLOVTKIKPMIIMTITTKDLAIAACVPRYSSTLEDRQPPASVYFNPDLITTW EDSMANVAGENIIYISRCHLNLSENTEPMDPTP" | complement(18848..19345) /note="ORF B04; short ORF of immediate early transcript 1 RNA" /codon_start=1 |
| gene | | |
| Query Match | 12.6%; Score 39.6; DB 14; Length 108873; | |
| Best Local Similarity | 50.4%; Pred. No. 9.6; | |

[illegible]

CDS
complement(join(25359..25414,25792)..26337,26389..26506,
26531..26633,26676..26793,26862..26943,26989..27235,
27279..27402))
/gene="Y116A8A.6"
/note="Predicted using GeneFinder
cDNA EST yk82d6.5 comes from this gene
cDNA EST yk290b7.3 comes from this gene
cDNA EST yk339e8.3 comes from this gene
cDNA EST yk471a12.3 comes from this gene
cDNA EST yk290b7.5 comes from this gene
cDNA EST yk339e8.5 comes from this gene
cDNA EST yk471a12.5 comes from this gene"
/codon_start=1
/protein_id="CAB55163.1"
/db_xref="GI:5832826"
/translation="MKPQWIGVIVFGIIVNCLISSOKTVKENERKLKSGCNFLPI
SINQKGFGFENRKYTYNNKSCGHLILAGKNSTRYATAPFISERHILLISRWL
TAERHWIMNGFRDKNCSGGRHLDVPDIVDNLFWGKKPKRADISKAMFPACN
ODLKPVALIELNKTSKRPMAPNVGPCADFTSSGLDVSHSGYGYESOHRKLSI
TAVKERICTPTYTEKAKGCPILINVSKAIVILKAPASRENCLEFYDSLQKEIC
DYSGAGEHNPLITTEPEPTTERSTVSTISTLETITMOASSASPISSKATTTP
SPLOASSPAEPILLOPESARTASASSARENOVTJPTETOPSLSKTSIKTSTYT
TTTYYTTTPDATVDYGDHDHMGVLVEDEDETITSNNRAHMIDLITGFEPFLSFET
EQCKI"

CDS
join(29644..29686,29735..30195)
/gene="Y116A8A.7"
join(29644..29686,29735..30195)
/gene="Y116A8A.7"
/note="contains similarity to Pfam domain: PF00635 (MSP
(Major sperm protein) domain), Score=61.1,
E-value=7.5e-15, N=1"
/codon_start=1
/protein_id="CAB55166.2"
/db_xref="GI:6425238"
/translation="MNNSLICRLRLFDFHDLSTKNNDSAFLVSLTKRNTIEYOS
VPPGVATIEPAKKIIIFNAFPDSMHYQVKALINRTIAYIRILNMKRSIYPCCI
LKTDQNMFEFIITFAFNYYTENTKMDRIITYKWINTPNNEDEYEFREWFGDGWVQH
IYOINI"

gene
join(31256..31363,31418..31531,31575..31728,31777..31947,
31990..32269,32332..32376)
/gene="Y116A8A.8"
join(31256..31363,31418..31531,31575..31728,31777..31947,
31990..32269,32332..32376)
/gene="Y116A8A.8"
/note="predicted using geneWISE
cDNA EST yk540c6.5 comes from this gene"
/codon_start=1
/protein_id="CAB55160.1"
/db_xref="GI:5832823"
/translation="MITHHRALMTCNSTEYSITEWRDCIELCAHNSCYAQTNLTV
RKECWCSTIONLMYIQRRFSANNMMNIIALKTDDTINETSEYCTPEEGIVGNMNTS
NLTOYKVAQPSBEDGNLFHFYSAIKSCPSEMPEFDRLRGCMCTKVFROGSDDHAR
ACKREANVALTRILETRKEMEYWMTAYVEALFKOPSMWEYLETWIDVGVRNCEOVN
WMSICGCDOGEDEFEDKISKGGYTMSDKNPRDLNWQCCLVIAMINPNKREVDDDA
CGSNKIKGPGOCREAG"

gene
complement(join((47919..48041,49593..50165,50969..51125,
51584..51908,52795..52899,54267..54318)))
/gene="Y116A8A.9"
complement(join(47919..48041,49593..50165,50969..51125,
51584..51908,52795..52899,54267..54318)))
/gene="Y116A8A.9"
/note="contains similarity to Pfam domain: PF00557
(metallopeptidase family M24), Score=239.2,
E-value=1.9e-68, N=1
cDNA EST EMBL:Z14814 comes from this gene
cDNA EST EMBL:Z14913 comes from this gene
cDNA EST EMBL:M89428 comes from this gene
cDNA EST yk138a12.5 comes from this gene
cDNA EST yk216d3.5 comes from this gene
cDNA EST yk284a3.5 comes from this gene

CDS
complement(join(25359..25414,25792)..26337,26389..26506,
26531..26633,26676..26793,26862..26943,26989..27235,
27279..27402))
/gene="Y116A8A.6"
/note="Predicted using GeneFinder
cDNA EST yk82d6.5 comes from this gene
cDNA EST yk290b7.3 comes from this gene
cDNA EST yk339e8.3 comes from this gene
cDNA EST yk471a12.3 comes from this gene
cDNA EST yk290b7.5 comes from this gene
cDNA EST yk339e8.5 comes from this gene
cDNA EST yk471a12.5 comes from this gene"
/codon_start=1
/protein_id="CAB55163.1"
/db_xref="GI:5832826"
/translation="MKPQWIGVIVFGIIVNCLISSOKTVKENERKLKSGCNFLPI
SINQKGFGFENRKYTYNNKSCGHLILAGKNSTRYATAPFISERHILLISRWL
TAERHWIMNGFRDKNCSGGRHLDVPDIVDNLFWGKKPKRADISKAMFPACN
ODLKPVALIELNKTSKRPMAPNVGPCADFTSSGLDVSHSGYGYESOHRKLSI
TAVKERICTPTYTEKAKGCPILINVSKAIVILKAPASRENCLEFYDSLQKEIC
DYSGAGEHNPLITTEPEPTTERSTVSTISTLETITMOASSASPISSKATTTP
SPLOASSPAEPILLOPESARTASASSARENOVTJPTETOPSLSKTSIKTSTYT
TTTYYTTTPDATVDYGDHDHMGVLVEDEDETITSNNRAHMIDLITGFEPFLSFET
EQCKI"

```

Query Match      12.5%: Score 39.4; DB 3; Length 54700;
Best Local Similarity 47.1%; Pred. No. 12;
Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY    12 tcagaattttaagaataacttttgagaaatgagacaatatgaatcagtlctccaccact 71
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    3851 TCATGCTTTTAAAGCAACATTTTTTTCCTAATTGATGAATGGCCTTGATATTTAAATAA 3792

QY    72 taagatatctcttagagatctacagcctccctttagggagacatacaaatgcagtgtgt 131
      || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    3791 TGTATATAGGATATATTTCTCCAGAACACAGATGAATTTCTTAATAACTTCACGCTTTCA 3732

QY    132 tgccttgttgagtcgccacctataltcaaglaagtlagctacaaaatlittgaaataga 191
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    3731 TCTTTGAGTCATTTTGGACATGCTTAAACAATCTTTTGAAGTCATATATGATTTTGTA 3672

QY    192 ttgtcacacaataaacctggaglttatggaaacatcagtagaaggaaatlacaaatlccat 251
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db    3671 AAGTCAAAAAAAACCTCAATTAATGATGATCTGATGCACACGMAATCTAAAGAAATG 3612

QY    252 cccttacagagatcat 268
      ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    3611 AAATTTCGAAAAATCAT 3595

RESULT 8
CEV113B8_1/c
WPCOMMENT
Sequence split into 4 fragments LOCUS CEV113B8 Accession AL021573
Fragment Name Begin End
CEV113B8_0 1 110000
CEV113B8_1 100001 210000
CEV113B8_2 200001 310000
CEV113B8_3 300001 357994
Continuation (2 of 4) of CEV113B8 from base 100001 (AL021573 Caenorhabditis elegans c

Query Match      12.5%: Score 39.4; DB 2; Length 110000;
Best Local Similarity 47.1%; Pred. No. 11;
Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY    12 tcagaattttaagaataacttttgagaaatgagacaataatgaatcagtlctccaccact 71
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    102826 TCATGCTTTTAAAGCAACATTTTTTTCCTAATTGATGAATGGCCTTGATATTTAAATAA 102767

QY    72 taagatatctcttagagatctacagcctccctttagggagacatacaaatcagtlgtgt 131
      || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    102766 TGTATATAGGATATATTTCTCCAGAACACAGATGAATTTCTAATAACTTCACGCTTTCA 102707

QY    132 tgccttgttgagtcgccacctataltcaaglaagtlagctacaaaatlittgaaataga 191
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    102706 TCTTTGAGTCATTTTGGACATGCTTAAACAATCTTTTGAAGTCATATATGATTTTGTA 102647

QY    192 ttgtcacacaataaacctggaglttatggaaacatcagtagaaggaaatlacaaatlccat 251
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db    102646 AAGTCAAAAAAAACCTCAATTAATGATGATCTGATGCACACGMAATCTFAAAGAANAATG 102587

QY    252 cccttacagagatcat 268
      ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    102586 AAATTTCGAAAAATCAT 102570

RESULT 9
CEV113B8_2/c
WPCOMMENT
Sequence split into 4 fragments LOCUS CEV113B8 Accession AL021573

```

Fragment Name Begin End
 CEY113B8_0 1 110000
 CEY113B8_1 100001 210000
 CEY113B8_2 200001 310000
 CEY113B8_3 300001 357994
 Continuation (3 of 4) of CEY113B8 from base 200001 (AI021573 Caenorhabditis elegans chrC)

Query Match 12.5%; Score 39.4; DB 2; Length 110000;
 Best Local Similarity 47.1%; Pred. No. 11;
 Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 12 tcaagatttaagaataaacttttgagaataagaacaaatgaatcaatttccaccact 71
 Db 2826 TCATGTTTTTAATGAACTTTTTCCTTAATGATGAATGGCCTTGATTTAAATAA 2767
 Qy 72 taagatactccttagagatctacagcctcctttagggagacacaaagtcagtgtg 131
 Db 2766 TGTATATGATATATTTATTCAGAAACACGATGAATTTTCTAATAAACCTTCACCTTTCA 2707
 Qy 132 tgccttggtagtccaccctatatatcaagtagtactacaattttgaaataga 191
 Db 2706 TCTTTGAGTCAATTTTGCACATGCTTAACATCTTTGAGTCAATATGATTTGTA 2647
 Qy 192 ttgtcaccaataaactcggagtttatggaaacatcagtagaaggaataacacattccat 251
 Db 2646 AAGTCAAAAAATTAACCTCAATTAATGATGATGATGACACGGAATCTTAAGAAATG 2587
 Qy 252 cccctacagagatcat 268
 Db 2586 AAATTTGCAAAATCAT 2570

RESULT 10
 AC092859/c 171179 bp DNA linear HTG 01-NOV-2001
 LOCUS Pan troglodytes clone rp43-11lm15, WORKING DRAFT SEQUENCE, 8
 DEFINITION unordered pieces.
 AC092859
 VERSION AC092859.9 GI:16519522
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE chimpanzee.
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 1 (bases 1 to 171179)
 N1.Y., Song,L. and Roe,B.A.
 Pan troglodytes BAC Clone rp43-11lm15
 Unpublished
 2 (bases 1 to 171179)
 N1.Y., Song,L. and Roe,B.A.
 Direct Submission
 Submitted (31-JUL-2001) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 On Oct 30, 2001 this sequence version replaced gi:16445160.
 ----- Genome Center
 Center: Department Of Chemistry And Biochemistry
 The University Of Oklahoma
 Center code:UOKNOR

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 2757 2756: contig of 2756 bp in length
 * 2857 2856: gap of unknown length
 * 5887 5886: contig of 3030 bp in length
 * 5887 5886: gap of unknown length

* 5987 12277: contig of 6291 bp in length
 * 12278 12377: gap of unknown length
 * 12378 20375: contig of 7998 bp in length
 * 20376 20475: gap of unknown length
 * 20476 37432: contig of 16957 bp in length
 * 37433 37532: gap of unknown length
 * 37533 60493: contig of 22960 bp in length
 * 60493 60592: gap of unknown length
 * 60593 108314: contig of 47721 bp in length
 * 108314 171179: gap of 62766 bp in length.
 * 108414 location/Qualifiers
 1..171179
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="rp43-11lm15"
 /clone_1lb="RPCT-43"

BASE COUNT 47434 a 39162 c 37372 g 46496 t 715 others
 ORIGIN

Query Match 12.5%; Score 39.4; DB 2; Length 171179;
 Best Local Similarity 46.5%; Pred. No. 9.9;
 Matches 127; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 9 gactcagatttaagaataaacttttgagaataagaacaaatgaatcaatttccacc 68
 Db 27477 GATTCAAGAGCTTCTGTAAGTAATGATGAATAAAGCAATTAATACCTTTGATGACC 27418
 Qy 69 acttaagatatactcttagagatctacagcctcctttagggagacataaagtcagtc 128
 Db 27417 ACTGTAGAAATTTAATAGCTATCATCACTTGCACATAAGATGTATTAAGGTAGGCA 27358
 Qy 129 ttttgccttggtagtccaccctatatatcaagtagtactacaattttgaaat 188
 Db 27357 AGAATCCTGAAGAGCTTTCTAAGCCCAAGTTCATTTGAAGAAATGAAGCAAGTTTCAAAAT 27298
 Qy 189 agattgtcacacaataaactcggagtttatggaaacatcagtagaaggaataacacattc 248
 Db 27297 ACAGTCATATGTCGATTAACATGTTTGTGCACACAGAGCGCATTAATGTATGCTAG 27238
 Qy 249 catcccttaacagagatcatattcattgcacattc 281
 Db 27237 TGATCCTGTAAAGATTATATGAGCTGAACAC 27205

RESULT 11
 HSJ364H10/c 150355 bp DNA linear PRI 21-FEB-2000
 LOCUS HSJ364H10
 DEFINITION Human DNA sequence from clone RP3-364H10 on chromosome 6q16.1-16.3
 ACCESSION AL078603
 VERSION AL078603.4 GI:6456827
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 150355)
 Williams,S.
 Direct Submission
 Submitted (21-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Nov 21, 1999 this sequence version replaced gi:5738453.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,

COMMENT

such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP3-364H10 is Institute by the library RP3-3 constructed at the Roswell Park Cancer

<http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP3-364H10. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true right end of clone RP3-364H10 is at 150355 in this

sequence. The true right end of clone RPI-273N12 is at 100 in this

sequence. This sequence was generated from part of bacterial clone configs of human chromosome 6, constructed by the Sanger Centre

<http://www.sanger.ac.uk/HGP/Chr6>.

FEATURES

source

Location/Qualifiers

1. 150355

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/map="q16.1-16.3"

/clone="RP3-364H10"

/clone_1lb="RP3-3"

3181. 3328

repeat_region /note="4 copies 12 mer 85% conserved"

3183. 3226

repeat_region /note="22 copies 2 mer ca 88% conserved"

7440. 7511

repeat_region /note="6 copies 12 mer 87% conserved"

7441. 7508

repeat_region /note="34 copies 2 mer ag 89% conserved"

complement(8514..9064)

/note="match: GSS: Em:AQ317776"

9398. 9517

repeat_region /note="10 copies 12 mer 64% conserved"

9431. 9515

repeat_region /note="17 copies 5 mer atata 65% conserved"

9875. 10098

repeat_region /note="HERVL repeat: matches 4335. 4548 of consensus"

10429. 10684

repeat_region /note="MER77 repeat: matches 81. 336 of consensus"

10740. 11170

repeat_region /note="MSTD repeat: matches 1. 394 of consensus"

15206. 15773

repeat_region /note="MER89 repeat: matches 1. 559 of consensus"

complement(16716..17027)

/note="match: STS: Em:HSPA32D12"

21089. 21629

misc_feature /note="match: STS: Em:Z78974"

21100. 21488

misc_feature /note="match: GSS: Em:AQ030832"

23485. 23606

repeat_region /note="61 copies 2 mer ac 64% conserved"

23490. 23609

repeat_region /note="10 copies 12 mer 69% conserved"

23972. 24389

misc_feature /note="match: GSS: Em:AQ240903"

24019. 24541

misc_feature /note="match: GSS: Em:AQ310408"

28622. 28851

misc_feature /note="match: GSS: Em:B75442"

misc_feature /note="match: GSS: Em:AQ279883"

29568. 30147

misc_feature /note="match: GSS: Em:AQ322861"

repeat_region

30996. 31342

misc_feature /note="MER58B repeat: matches 2. 340 of consensus"

complement(40152..40587)

/note="match: GSS: Em:AQ17754"

misc_feature /note="match: GSS: Em:AQ014036; match: STS: Em:G51620"

complement(40213..40811)

/note="match: GSS: Em:AQ014036; match: STS: Em:G51620"

complement(40357..40781)

/note="match: GSS: Em:B86714"

40819. 41341

misc_feature /note="match: GSS: Em:AQ211747"

46207. 46457

misc_feature /note="match: GSS: Em:AQ307766"

49669. 49929

repeat_region /note="match: GSS: Em:AQ307766"

52044. 52403

misc_feature /note="match: GSS: Em:AQ165687"

53699. 54781

repeat_region /note="TIGER3B repeat: matches 1. 1054 of consensus"

54775. 54811

repeat_region /note="TIGER3(Golem) repeat: matches 2861. 2897 of

consensus"

54819. 54966

repeat_region /note="TIGER3(Golem) repeat: matches 2298. 2456 of

consensus"

54957. 55129

repeat_region /note="TIGER3B repeat: matches 1069. 1241 of consensus"

61258. 61598

repeat_region /note="LTR18A repeat: matches 1. 346 of consensus"

61600. 62465

repeat_region /note="HERVL18 repeat: matches 5189. 6047 of consensus"

62466. 62921

repeat_region /note="MTR2B repeat: matches 4. 444 of consensus"

62922. 68052

repeat_region /note="HERVL18 repeat: matches 5. 5189 of consensus"

68053. 68404

repeat_region /note="LTR18A repeat: matches 1. 356 of consensus"

complement(70793..71290)

/note="match: GSS: Em:AQ008841"

complement(71012..71269)

/note="match: GSS: Em:B83962"

complement(74346..74831)

/note="match: STS: Em:Z79249"

75056. 75504

repeat_region /note="L1MC1 repeat: matches 5653. 6108 of consensus"

75929. 76148

repeat_region /note="L1MC1 repeat: matches 6104. 6327 of consensus"

78209. 78430

repeat_region /note="LTR18B repeat: matches 218. 462 of consensus"

78873. 79225

repeat_region /note="MER58B repeat: matches 1. 337 of consensus"

84416. 86511

repeat_region /note="L1P repeat: matches 1601. 3701 of consensus"

complement(86523..86971)

/note="match: GSS: Em:B88696"

88350. 88397

repeat_region /note="4 copies 12 mer 100% conserved"

92738. 93167

repeat_region /note="MTR2B repeat: matches 2. 448 of consensus"

93344. 93385

repeat_region /note="MTR2B repeat: matches 459. 504 of consensus"

97891. 98411

repeat_region /note="MTR2B repeat: matches 1. 448 of consensus"

100416. 100639

repeat_region /note="MER63A repeat: matches 2. 206 of consensus"

complement(101069..101516)

/note="match: GSS: Em:AQ211414"

102301. 103031

misc_feature /note="match: GSS: Em:AQ027024"

complement(109215..109848)

/note="match: GSS: Em:B56601"

complement(109418..109858)

/note="match: GSS: Em:AQ217743"

complement(109446..109864)

```

misc_feature      /note="match: GSS: Em:B53315"
                  complement(109700..109858)
repeat_region     /note="match: STS: Em:284791"
                  110200..110661
repeat_region     /note="TIGER2 repeat: matches 3..2718 of consensus"
                  112200..112686
repeat_region     /note="HERV16 repeat: matches 1929..2435 of consensus"
                  112832..112832
repeat_region     /note="MER93 repeat: matches 1..83 of consensus"
                  113518..113936
                  /note="Tiger3(Golem) repeat: matches 2597..3027 of
                  consensus"
repeat_region     113960..114079
                  /note="HERV1 repeat: matches 5226..5344 of consensus"
repeat_region     114199..114633
                  /note="LTR16B repeat: matches 5..464 of consensus"
misc_feature      /note="complement(116628..116977)
                  /note="match: GSS: Em:A0310407"
                  complement(116724..116977)
misc_feature      /note="match: GSS: Em:A0240902"
                  /note="match: GSS: Em:A0240902"
repeat_region     120407..120634
                  /note="19 copies 12 mer 75% conserved"
repeat_region     120408..120629
                  /note="111 copies 2 mer ga 64% conserved"
repeat_region     121253..121324
                  /note="36 copies 2 mer tc 84% conserved"
repeat_region     123359..123400
                  /note="21 copies 2 mer ct 81% conserved"
repeat_region     123140..123710

```

```

Query Match      12.4%; Score 39; DB 9; Length 150355;
Best Local Similarity 50.8%; Pred. No. 13;
Matches 93; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```

```

QY 69 acttaagatatactccttagagatacagcctccttagaggagacatacaagaagtcagt 128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101572 ACTTGACCTAAATTCCTTTAAACTCTACTCTCTCCAGGTAAATCTGAATTCGTA 101513

QY 129 ttgtcccttgtagagccaccctatactcaagtagtagtaactacaatttgaagaat 188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101512 TGTACTCTCTCTTTAGGTACATTAATGTCACGATATTTATTAATAATATAGGAA 101453

QY 189 agattgtcacacaataactgaggttattggaacatcagtagaagaagaatacaacattc 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101452 ATGTTGTCACACTTAAGACCAAAATTTGTGATTCATGCTGCAAGCATATACATTA 101393

QY 249 cat 251
    ||
Db 101392 AAT 101390

```

```

RESULT 12
AC011935      AC011935      168043 bp      DNA      linear      HTG 20-SEP-2000
DEFINITION   Homo sapiens clone RP11-16C5, WORKING DRAFT SEQUENCE, 17 unordered
              pieces.
ACCESSION    AC011935
VERSION      AC011935.4      GI:10198441
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 168043)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
              Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
              Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
              Cooke,P., Deailellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
              Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
              JOURNAL
              Unpublished
              2 (bases 1 to 168043)
              Homo sapiens, clone RP11-16C5
              and Lander,E.

```

TITLE JOURNAL COMMENT

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lebecky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwen,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliou,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:9123857.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information
Center project name: L3518
Center clone name: 16_C-5

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: PHRAP; version 0.960731
Consensus quality: 159963 bases at least Q40
Consensus quality: 163758 bases at least Q30
Consensus quality: 165359 bases at least Q20
Insert size: 166443; sum-of-ctrls
Quality coverage: 4.8 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1
2726 2825: contig of 2725 bp in length
2826 4365: contig of 1540 bp in length
4366 4465: gap of 100 bp
4466 6534: contig of 2069 bp in length
6535 6634: gap of 100 bp
6635 8203: contig of 1569 bp in length
8204 8303: gap of 100 bp
8304 34742: contig of 26439 bp in length
34743 34842: gap of 100 bp
34843 39378: contig of 4536 bp in length
39379 39478: gap of 100 bp
39479 45278: contig of 5800 bp in length
45279 45378: gap of 100 bp
45379 49598: contig of 4420 bp in length
49599 49698: gap of 100 bp
49699 56999: contig of 7301 bp in length
57000 57099: gap of 100 bp
57100 62962: contig of 5863 bp in length
62963 63062: gap of 100 bp
63063 74352: contig of 11290 bp in length
74353 74452: gap of 100 bp
74453 88200: contig of 13748 bp in length
88201 88300: gap of 100 bp
88301 104524: contig of 16224 bp in length
104525 104624: gap of 100 bp
104625 123085: contig of 18461 bp in length
123086 123185: gap of 100 bp
123186 141113: contig of 17928 bp in length
141114 141213: gap of 100 bp
141214 166727: contig of 25514 bp in length

```

| FEATURES | * | 166728 | 166827 | gap of | 100 bp |
|--|---|-------------------------------------|--------|------------------------------|--------|
| SOURCE | * | 166828 | 168043 | contig of 1216 bp in length. | |
| | | Location/Qualifiers | | | |
| | | 1. 168043 | | | |
| | | /organism="Homo sapiens" | | | |
| | | /db_xref="taxon:9606" | | | |
| | | /clone="RP11-16C5" | | | |
| | | /clone_lib="RPC1-11 Human Male BAC" | | | |
| | | 1. 2725 | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | clone_end:SP6 | | | | |
| | vector_side:left" | | | | |
| | 2826. 4365 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 4466. 6534 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 6635. 8203 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 8304. 34742 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 34843. 39378 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 39479. 45278 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 45379. 49598 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 49699. 56999 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 57100. 62962 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 63063. 74352 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 74453. 88200 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 88301. 104524 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 104625. 123085 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 123186. 141113 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 141214. 166727 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | 166828. 168043 | | | | |
| misc_feature | /note="assembly_fragment" | | | | |
| | clone_end:T7 | | | | |
| | vector_side:right" | | | | |
| BASE COUNT | 49133 a 33769 c 33077 g 50464 t 1600 others | | | | |
| ORIGIN | | | | | |
| Query Match | 12.4% Score 39; DB 2; Length 168043; | | | | |
| Best Local Similarity | 50.8%; Pred. No. 12; | | | | |
| Matches 93; Conservative 0; Mismatches 90; Indels 0; Gaps 0; | | | | | |
| 69 | acttaagratctctcttgagatctcaagccctcttagggagacatacaagtcagtg 128 | | | | |
| 20556 | ACTTGACCAAAATTTCTTTAAAACTTCTTACTTCTCTCGGTAAATCTGAATTCGTGA 20615 | | | | |
| 129 | tgttcgcttglttgagtcaccactatataccaagtaggtatgactacaatttgaanaat 188 | | | | |
| 20616 | TGTACTCTTCTTTAGTAGTACATFAAAGTGCACAGATATTTTATTAATAATAAGGAA 20675 | | | | |
| 189 | agattgtaccacaataaactggagcttcttggaacaatcagtcagaaggaatacaacatc 248 | | | | |
| 20676 | ATGTTGTCAACTTAAGACCAAAAAATTTGTATTCATGTGTGAAGGATTAAGAAATATTA 20735 | | | | |
| 249 | cat 251 | | | | |
| 20736 | AAT 20738 | | | | |

| | | | | | |
|-----------------------|---|-----------------|-----------|----------------|-----------------|
| LOCUS | AC013287 | 174576 bp | DNA | linear | PRI 10-AUG-2001 |
| DEFINITION | Homo sapiens chromosome 10 clone RP11-76F14, complete sequence. | | | | |
| ACCESSION | AC013287 | | | | |
| VERSION | AC013287.10 | GI:15147146 | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | | | |
| AUTHORS | Smith,D.R. | | | | |
| TITLE | Genome Therapeutics Corporation Sequencing Center: Human Genome Sequence Data | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 174576) | | | | |
| AUTHORS | Smith,D.R. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (06-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA | | | | |
| REFERENCE | 3 (bases 1 to 174576) | | | | |
| AUTHORS | Smith,D.R. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (10-AGC-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA | | | | |
| COMMENT | On Aug 10, 2001 this sequence version replaced gi:15011672. | | | | |
| FEATURES | Location/Qualifiers | | | | |
| SOURCE | 1. 174576 | | | | |
| | /organism="Homo sapiens" | | | | |
| | /db_xref="taxon:9606" | | | | |
| | /chromosome="10" | | | | |
| | /clone="RP11-76F14" | | | | |
| | /clone_1fb="RPC1-11" | | | | |
| BASE COUNT | 53503 a 30664 c 31890 g 58519 t | | | | |
| ORIGIN | | | | | |
| Query Match | 12.4%; | Score 39; | DB 9; | Length 174576; | |
| Best Local Similarity | 45.8%; | Pred. No. 12; | | | |
| Matches 135; | Conservative 0; | Mismatches 160; | Indels 0; | Gaps 0; | |
| Y | 21 aagaatacacttttggagaatagagaacaatgaataacagttctccaccacttaagtatat 80 | | | | |
| | | | | | |
| Db | 7027 AAGAAAGTCTCTTTGGCTTATGAAAAGTCATAAATTTTATTTCTCTAAATTTATGTTAGG 7086 | | | | |
| Y | 81 ccttagagagactcagcgcctcccttagaggagacatacaagaatcagttgltgccttgt 140 | | | | |
| | | | | | |
| Db | 7087 ATAAAGACAATATCTGCAATTTATGTAATACACATTAATGTCCTTTGGCAGAGTTTGT 7146 | | | | |
| Y | 141 tgaatcccaacttatattcaagttagttagtactacaatttgaanaatagattgccaac 200 | | | | |
| | | | | | |
| Db | 7147 TAAATTTAAACAAGCATTTTCTATACCTCATTTATTCATTGGTTTAAATTACCTTTATGAA 7206 | | | | |
| Y | 201 aataaactggaagttatggaacaactcagtagaagaagaatacaacaatccatcccttaca 260 | | | | |
| | | | | | |
| Db | 7207 GAATATATATTTAAAGTATATTTATAGTTCGAGTAAGACAAATTAATTAATTCATACCA 7266 | | | | |
| Y | 261 gagatcatttccttgcgaactcagagataaattgtcattcgtattatcactcatgc 315 | | | | |
| | | | | | |
| Db | 7267 TTTTGGTTAACTTTGATCCCTTGAATATTTTGCACTTATTTAAATATATTTTTTAC 7321 | | | | |
| RESULT 14 | | | | | |
| LOCUS | AC012075 | 185437 bp | DNA | linear | PRI 20-APR-2001 |
| DEFINITION | Homo sapiens clone RP11-492P1, complete sequence. | | | | |
| ACCESSION | AC012075 | | | | |
| VERSION | AC012075.10 | GI:13677160 | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 185437) | | | | |

[illegible]

Search completed: September 16, 2002, 02:57:19
Job time: 7652 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 00:49:47 : Search time 3785.3 Seconds
(without alignments)
9287.661 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 gctgctgagagataagtt.....tcaataaacactgctgtg 1680

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vl:*
30: em.htg.hum:*
31: em.htg.in:*
32: em.htg.other:*
33: em.htgo.in:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|-------|--------|----|----|-------------|
|------------|-------|-------|-------|--------|----|----|-------------|

| | | | | | | |
|------|--------|-------|--------|----|-----------|---------------------|
| 1 | 1679.2 | 100.0 | 1680 | 6 | AX224643 | AX224643 Sequence |
| 2 | 1656.2 | 98.6 | 2036 | 6 | AX224647 | AX224647 Sequence |
| 3 | 1617 | 96.2 | 1636 | 6 | AX092346 | AX092346 Sequence |
| 4 | 1615.8 | 96.2 | 1684 | 9 | BC008429 | BC008429 Homo sapi |
| 5 | 822.8 | 49.0 | 1629 | 10 | MMU6068 | U6068 Mus musculu |
| 6 | 806 | 48.0 | 1035 | 6 | AX301908 | AX301908 Sequence |
| 7 | 724.2 | 43.1 | 963 | 6 | AX301906 | AX301906 Sequence |
| 8 | 181.4 | 10.8 | 159440 | 9 | AL355392 | AL355392 Human DNA |
| 9 | 161 | 9.6 | 161593 | 2 | HSRA94G10 | AL121901 Human DNA |
| c 10 | 83.2 | 5.0 | 89536 | 2 | AC094188 | AC094188 Rattus no |
| c 11 | 80.6 | 4.8 | 92996 | 2 | AC094264 | AC094264 Rattus no |
| 12 | 77.6 | 4.6 | 68591 | 2 | AC107771 | AC107771 Mus muscu |
| 13 | 59.4 | 3.5 | 89536 | 2 | AC094188 | AC094188 Rattus no |
| 14 | 58.8 | 3.5 | 125020 | 9 | AF429315 | AF429315 Homo sapi |
| 15 | 54.6 | 3.2 | 999 | 4 | BTU79413 | AF429315 Homo sapi |
| 16 | 54 | 3.2 | 92996 | 2 | AC094264 | U79413 Bos taurus |
| 17 | 50.4 | 3.0 | 1011 | 6 | AF172993 | AC094264 Rattus no |
| 18 | 48.8 | 2.9 | 768 | 6 | E53840 | AF172993 Homo sapi |
| 19 | 48.8 | 2.9 | 1015 | 6 | E53841 | E53840 LUNX gene a |
| 20 | 48.8 | 2.9 | 1015 | 6 | AB024937 | E53841 LUNX gene a |
| 21 | 48.8 | 2.9 | 1053 | 9 | AB024937 | AB024937 Homo sapi |
| 22 | 48.8 | 2.9 | 1084 | 9 | AF158745 | BC012549 Homo sapi |
| 23 | 47.8 | 2.8 | 33285 | 1 | SC5F8 | AF158745 Homo sapi |
| 24 | 47 | 2.8 | 5676 | 9 | AB033050 | AL357613 Streptomy |
| c 25 | 46.8 | 2.8 | 185005 | 2 | AC094500 | AB033050 Homo sapi |
| 26 | 46 | 2.7 | 1696 | 4 | AB055894 | AC094500 Rattus no |
| 27 | 46 | 2.7 | 31382 | 1 | SC3C3 | AB055894 Bos tauru |
| 28 | 45.2 | 2.7 | 1500 | 1 | AVSUCCTR | AL031231 Streptomy |
| 29 | 45 | 2.7 | 3312 | 1 | AF061246 | X52432 A.vinelandi |
| 30 | 44.8 | 2.7 | 1561 | 10 | AF194029 | AF061246 Ralstonia |
| c 31 | 44.8 | 2.7 | 125020 | 9 | AF429315 | AF194029 Rattus no |
| c 32 | 44.6 | 2.7 | 1692 | 3 | AF350278 | AF429315 Homo sapi |
| c 33 | 44.4 | 2.6 | 45335 | 9 | AF369213 | AF350278 Nephtila m |
| c 34 | 44.2 | 2.6 | 68591 | 2 | AC107771 | AF369213 Homo sapi |
| c 35 | 43.8 | 2.6 | 1675 | 10 | AB055895 | AL136319 Streptomy |
| 36 | 43.8 | 2.6 | 1713 | 10 | AB055895 | AC107771 Mus muscu |
| c 37 | 43.8 | 2.6 | 12213 | 1 | AE004441 | AB055895 Rattus no |
| c 38 | 43.8 | 2.6 | 20102 | 1 | SC9B2 | AB055895 Rattus no |
| 39 | 43.4 | 2.6 | 174584 | 2 | AC068495 | AE004441 Pseudomon |
| c 40 | 42.8 | 2.5 | 1668 | 10 | AF194028 | AL055212 Streptomy |
| 41 | 42.8 | 2.5 | 1691 | 10 | BC011168 | AC068495 Mus muscu |
| 42 | 42.8 | 2.5 | 1709 | 10 | AB055897 | AF194028 Mus muscu |
| 43 | 42.8 | 2.5 | 13591 | 10 | AB055898 | BC011168 Mus muscu |
| 44 | 42.8 | 2.5 | 13591 | 10 | AB055898 | AB055897 Mus muscu |
| 45 | 42.6 | 2.5 | 1584 | 5 | GGRRG2 | AB055898 Mus muscu |

ALIGNMENTS

RESULT 1
AX224643
LOCUS AX224643 1680 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 1 from Patent WO0161055.
ACCESSION AX224643
VERSION AX224643.1 GI:15554772
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
1 (bases 1 to 1680)
Chen,S.Y., Sun,Y. and Macina,R.A.
METHODS for diagnosing, monitoring, staging, imaging and treating
lung cancer via lung cancer specific genes
Patent: WO 0161055-A 1 23-AUG-2001;
diabex, Inc. (US)
JOURNAL
diabex, Inc. (US)
FEATURES
Source location/Qualifiers
1..1680
/organism="Homo sapiens"
/db_xref="taxon:9606"

| BASE COUNT | 380 a | 503 c | 429 g | 367 t | 1 others |
|------------|-------|-------|-------|-------|----------|
|------------|-------|-------|-------|-------|----------|

QY 1021 aagaattcaatgctcttggactctgtcttcctagatgccatgcyctgaagctaa 1080
|||||
Db 1021 aagaattcatgctcttggactctgtcttcctagatgccatgcyctgaagctaa 1080

| Query Match | 98.6% | Score 1656.2 | DB 6 | Length 2036 |
|-----------------------|--|--------------------------|------|-------------|
| Best Local Similarity | 99.7% | Prod No. 0 | | |
| QY 1021 | aaagattatcgtctctgttggactctgtgtcttcctc | gagatgtccatcgtcgtgaatcaa | 1080 | |
| Db 1021 | AAGAAATCATGTGCTCTGTGGACCTGTGCCTCTCGAGAGTGGCCCATCGCTGAAATCA | 1080 | | |
| QY 1081 | gcacfcggctcatalgaanaagctcgaataagctcggagctccaccgaatcgtgaga | 1140 | | |
| Db 1081 | GCATCGGGCTGATCAATGAAAGAGCTGCAGATTAAGCTTGGATCTTACCCAGATCGTGAAGA | 1140 | | |
| QY 1141 | tcctcaactcaggacactcccgagitttclatagaaccaagccatgccaaggtggcccaac | 1200 | | |
| Db 1141 | TCCTCAACTCAGGACACACCCGAGTTTATTATAGACCAAGGCCATGACCAAGGTGGCCCAAC | 1200 | | |
| QY 1201 | tgatcgcgcgtggaagtgttccccccgaatgaagccctccgcccctgtttacacctggca | 1260 | | |
| Db 1201 | TGATCGTGCTGGGAAGTGTTCCTCCCGATGAAGCCCTCCGCCCTTGTATACCTGGGCA | 1260 | | |
| QY 1261 | tcgaagccagctcgggaagctcagttttacaccaaaaggtgaccaactatatacctaactga | 1320 | | |
| Db 1261 | TCGAAGCCAGCTCGGAAGCTCAGTTTATACACCAAAAGGTGACCAACTTATACCTCACTTGA | 1320 | | |
| QY 1321 | ataacatcagctctgatacgtatcgaatcgaatgtgaactctggatctgtgttccaaactg | 1380 | | |
| Db 1321 | ATAACATCAGCTCTGATCGGATTCACAGTGTGAATCTTGGATTGGCTGGTTCACACTTG | 1380 | | |
| QY 1381 | atgttctgaaacaatcatcatcactgaatcatccactccactcctcgtcgtccgaacagaatg | 1440 | | |
| Db 1381 | ATGTCTGAAACAAATCATCATCTACGATCATCATCATCATCATCTCGTGGCGAACCAAGATG | 1440 | | |
| QY 1441 | gcaaataaagatctgggtgtcccaagtctcatctgtgtgaagccctggatctcggagcagctg | 1500 | | |
| Db 1441 | GCAAAATTAAGATCTGGGGTCCACGTGTCTATTGGTGAAGGCTTGGGATTCGAGGACGCTG | 1500 | | |
| QY 1501 | agctctcactgaccaaagatgcccctgtgtcttaactccaaactcctcctgtgtgaaacccaagt | 1560 | | |
| Db 1501 | AGCTCTCACTGACCAAGATGCCCCTGTGCTTACTCTACCTCACCTTCCTGTGGAAACCCAGCT | 1560 | | |
| QY 1561 | ctctcgtctcccaatgaagaacttgatgtgcaagcatcagggaagctgtgttcccaagctg | 1620 | | |
| Db 1561 | CTCTCTCTCTCCAGTGAACACTTGGATGCGACGACATCAAGGAAGCTGTGGTCCACACTTG | 1620 | | |
| QY 1621 | gaagatgggtgtgagactctatagacacatccctcctcgtcaataataaacaactgtcgtg | 1680 | | |
| Db 1621 | GAGTATGGGTGTGAGGCTCTATAGACCATCTCCTCTGCAATCAATTAACACTTGGCTGTG | 1680 | | |
| RESULT 2 | | | | |
| AX224647 | | 2036 bp | DNA | linear |
| LOCUS | AX224647 | | | |
| DEFINITION | Sequence 5 from Patent WO0161055. | | | |
| ACCESSION | AX224647 | | | |
| VERSION | AX224647.1 | | | |
| KEYWORDS | GI:15554774 | | | |
| SOURCE | human. | | | |
| ORGANISM | Homo sapiens | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| AUTHORS | 1 (bases 1 to 2036) | | | |
| TITLE | Chen,S.Y., Sun,Y., and MacIina,R.A. | | | |
| JOURNAL | Methods for diagnosis, monitoring, staging, imaging and treating lung cancer via lung cancer specific genes | | | |
| FEATURES | Patent: WO 0161055-A 5 23-AUG-2001; | | | |
| SOURCE | diadexus, Inc. (US) | | | |
| ORIGIN | Location/Qualifiers | | | |
| | 1..2036 | | | |
| | /organism="Homo sapiens" | | | |
| | /db_xref="taxon:9606" | | | |
| | 477 a 586 c 528 g 444 t 1 others | | | |

Matches 1678: Conservative 2: Mismatches 0: Indels 3: Gaps 2:

```
QY 1 ggtgtcagagataaagtgtgacttcacgaccactgcccggagag--gagrgagcg 58
    |||
Db 38 GGTGTCCAGAGATTAAGTTGACTTCCAGACCACACTGCCCGGAGAGAGAGAGAGCG 97
QY 59 ggcgagagatccagcggtgcccaggtctgcatctctgcatctgtgccccttgacactg 118
    |||
Db 98 GGCCTGAGACTCCAGAGTCCAGGCTGTGCATCTCGACCTTGTCTCCCTGTGACACGTG 157
QY 119 ggaagatgtgcgcccgtgtgaccttcaacctctctgtgtgttctgtgcagccacttga 178
    |||
Db 158 GGAAGATGGCCGCGCGTGAGACCTTCCACCTTCTGTGTGTGGCGGACGCCACTTTTA 217
QY 179 tcgaagccaccttgaagtcacatctgtgacttcaactctgtgcccgaagatcgaaga 238
    |||
Db 218 TCCAAAGCCACCTCATGCTCCACGCTGCACTCTCACTCGGCCAAAAGTATCAAGAA 277
QY 239 agctgacacagagagcttgaaagacacaaacgacacagacatctctgacagctgtgcgtgc 298
    |||
Db 278 ACTGTACACAGAGAGCTGGAAGACACACAGCCACGACTCTGCGAGCTGCCCTGC 337
QY 299 tcaatgtcactgtcggaagaaagcagccgagag--atccctgtgtgtggcagcctgtgtgaa 357
    |||
Db 338 TCAGTGTCCATGCGGGGAAAAAGCCAGCCGAGGACATCCCTGTGTGGGACAGCTGTGTGA 397
QY 358 cgtctctgaagacatcatctgtgtgaagtatcaagctaacatctccctcagctgtcag 417
    |||
Db 398 CGGTCTGTGAAGACATCATCTGTGCTGAAGGTATCATACACTAATCTCCAGCTGCAAG 457
QY 418 tgaagccctctgacaaatgacagagagctgtgactgaagaatcccccctgtgacatgtgtgc 477
    |||
Db 458 TGAAGCCCTCGGCAATGACACAGAGCTCTGTAGTCAAGTCCCTCGGACATGTGTGCTG 517
QY 478 gattcaacacgcccctgtgtcaagaacacatcggtgtgagttccacatgacgactgtgagccaa 537
    |||
Db 518 GATTCAACACGCCCCCTGTGTGCAAGACCATGTGTGAGTTCCACATGACGACTGTGAGGCCA 577
QY 538 ccaccaatccgcatgtgacacagatgtgcaagtgtgccccccgctgtgtccctcaatgtgactgt 597
    |||
Db 578 CCACCATCCGCAATGGACACCACTGCAAGTGGCCCCACCGCTGTGTCTCACTGACTGTG 637
QY 598 ccaccaacatgtgagacgtgtgcacatccaaatgtgtgatalaagctctctctctgtgtgaa 657
    |||
Db 638 CCACGAGCCATGGAGACCTGCGCATCTGCTGATAGCTCTCTCTCTCTCTCTCTCTCTCT 697
QY 658 ccttgaagcagatgtcatgaaacttccatgtgcatccctgcgccaacttaagtga 717
    |||
Db 698 CCTTAGCTAAGCAGGTGATGAACCTTCCTAGTGCATCTCTGCCCAATTCAGTGA 757
QY 718 agctgtgtccgctgtgtgtgaggttcccttaagtgtatgtacagacccctcgtgcagctgt 777
    |||
Db 758 AGCTGTGTCCCGTGTGATGAGGCTCTCTTAATGTGATGTATGACAGCTCTGTGACACTG 817
QY 778 tgaagttgcacatctccctcagacatgtgacgctgtgagttgacacctctgtatccgtgca 837
    |||
Db 818 TGAAGTGTGCCATTTCCCTCAGCATGTGACGCTGTGAGTTGACCTTCTGTATCTCGCA 877
QY 838 tcaaggtgtgacacacatctgactctacgtgtggggtgccaagttgtgtgacatcacagggaa 897
    |||
Db 878 TCAAGGTGTGACACATCTGACTCTACCTGGGGGCCAAGTGTGTGACATCACAGGGA 937
QY 898 tgaacaaagtgttcaataactctgtgacgttccctcgaatgtgcccacccgtgga 957
    |||
Db 938 TGAACAAAGTGTGATTAATCTGTGCAAGCTTCCCTGACAAATGCCACCTGTGACAAACATCC 997
QY 958 cgttcaagcctcatcgtgtgactgtgagagctgtgaaagctgcaatgtgtgtgtgtgtctc 1017
    |||
Db 998 CATTCAAGCCTCATCTGTGATGTGACAGAGTGTGTGAAGAGCTGTGAGTGTGTGTCTCTC 1057
QY 1018 caagaagaatcactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1077
    |||
Db 1058 CAGAAGAAATTCATGTCTGTGTGACTGTGTGTCTCTGTGAGAGTCCCATCGGCTGAAGT 1117
```

```
QY 1078 caagcatcggtgtgatcaatga 1137
    |||
Db 1118 CAAGCATCGGCGGTGATCATGAA 1177
QY 1138 agatccatacgaagacactccgag 1197
    |||
Db 1178 AGATCTTAATCAAGACACTCCGAG 1237
QY 1198 aactgtatgtgtgtgtgtgtgtgt 1257
    |||
Db 1238 AACTGTATCTGTGTGTGTGTGTGT 1297
QY 1258 gcatcgaaagcagctcggaagct 1317
    |||
Db 1298 GCATCGAAGCCAGCTGTGAAAGCT 1357
QY 1318 tgaataacatcagctctgtatcga 1377
    |||
Db 1358 TGAATTAACATCAAGCTGTGATCG 1417
QY 1378 ctgagtgtctgtgaaacatcatc 1437
    |||
Db 1418 CTGATGTCTGTGAAAAATCATCA 1477
QY 1438 atggcaaatgaagatctgtgtgtgt 1497
    |||
Db 1478 ATGGCAAAATTAAGATGTGGGTG 1537
QY 1498 ctgagtcctcactgacaaagatgt 1557
    |||
Db 1538 CTGAGTCTCTACATGACCAAGAT 1597
QY 1558 gctctcgtctctccaggtgaag 1617
    |||
Db 1598 GCTCTCTCTCTCTCTCTCTCTCT 1657
QY 1618 tggagatgtgtgtgtgtgtgtgt 1677
    |||
Db 1658 TGGAGATGTGGGTGTGAGCTTA 1717
QY 1678 gtg 1680
    |||
Db 1718 GTG 1720
```

RESULT 3
AX092346 1636 bp DNA linear PAT 21-MAR-2001
LOCUS Sequence 77 from Patent W00116318.
DEFINITION AX092346
ACCESSION AX092346.1 GI:13444487
VERSION
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1636)
Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
Wood,M.I.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 77 08-MAR-2001;
Genentech, Inc. (US)
FEATURES
source Location/Qualifiers
1..1636
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 374 a 492 c 411 g 359 t
ORIGIN

| Query Match | 96.2% | Score 1617 | DB 6 | Length 1636 |
|-----------------------|--------------|---|------------|-------------|
| Best Local Similarity | 99.88 | Pred. No. 0 | | |
| Matches 1628 | Conservative | 1 | Mismatches | 1 |
| | | | Indels | 1 |
| | | | Gaps | 1 |
| QY | 51 | grrgagcgggacggagactccagcagctgcccagctctgtgcactcctgtcactgtgcctct | 110 | |
| Db | 1 | GAGAGAGGGGCGCGGAGACTCCAGCGTCCAGGTCTGTGGATCTCTGCATCTGCTGCCCTCT | 60 | |
| QY | 111 | gaaccctgggaaagatggccggccgtggagacttcaacctctctgtgtttgtgtgcaac | 170 | |
| Db | 61 | GACACCTGGGAAGATGGCCG6CCGtggACTTCACCTTCCTGTGGTTCTGTGCAAC | 120 | |
| QY | 171 | caacctgtaccaaagccacctcagtcctcactgtcagttctcactcactcctgtgcgcaaaagt | 230 | |
| Db | 121 | CACCTTGATCCAAAGCCACCCTCAGTCCACTGCAGTTTCATCTCTGGCCCAAAATCAT | 180 | |
| QY | 231 | caaaagaaagtctacacaggaagtgtgaagaccacaacgccacagaactctgtcaagagct | 290 | |
| Db | 181 | CAAAAGAAAGCTCACACAGGAGGTGAAGGACCAACAGCCACAGCATCTCAGACAGCT | 240 | |
| QY | 291 | gccgctgtactagtcgcatctgsgaaagaacgacggaggt-atccctgtgtgtggaagct | 349 | |
| Db | 241 | GCCGCTGCTAGTGGCCATGGGGAAGGCAAGCCGAGGAGCATCCTGTGTGTGGAGGCT | 300 | |
| QY | 350 | grrgaacacccgtctcgaagcaactactgtgtgaaagtctacagaagctaaactctcca | 409 | |
| Db | 301 | GGGGAACCCGTCCTTAAGCACAATCATCTGGGTGAAGGTCAACAGTAACATCTCTCA | 360 | |
| QY | 410 | gctgcaggtgaagccctctgcacatbaacaggaagctgtcagtcgaagatccccctggaat | 469 | |
| Db | 361 | GCTGCAAGTGAAGCCCTCGGCCAATGACAGAGAGCTGTAGTCAAGATCCCCCTGGACAT | 420 | |
| QY | 470 | grrgtgtgtattcaaaagccccctgttgaagaacactcgtlbgagtctcaatgacagtca | 529 | |
| Db | 421 | GGGTGCTGATTCACACCCCTCCCTGGTCAAGACCATCTGGAGTTCCACATGACACTGA | 480 | |
| QY | 530 | ggcccaagccacacatcccgatgagaccaaagctgaagttggccccaccgcccgttccatg | 589 | |
| Db | 481 | GGCCCAAGCCACCATCCGATGGACACAGTGCAGTGGCCCCACCCGCTGTGCTCAG | 540 | |
| QY | 590 | tgaactgtccaccagccatggagagccttgcgaatccaaactgtgtcaataagctctctcct | 649 | |
| Db | 541 | TGACTGTGCACACCAACCATGGGAGCCTCGCATTCACATGCTGTATAGTCTCTCTCCT | 600 | |
| QY | 650 | ggltgaacgcttaagtaagcaaggtatgaactcctctatgtccatccctgcgcccaatcagt | 709 | |
| Db | 601 | GGTGAACGCTTAGTGAAGCAGGTATATACCTCTTATGTGCCATCTCCGCCAATCTAGT | 660 | |
| QY | 710 | gaaaaacaaagctgtgtcccgltgacgaagcttctctcaatlgacatgtatgacaaacctct | 769 | |
| Db | 661 | GA AAAACCAAGCTGTCTCCCGTATGAGGCTTCTTCAATGACATGATAGAGACCTCCT | 720 | |
| QY | 770 | gaagctgtgtgaaggtggccattctccctcgaagtattgacccgttctgtgaagtttgaactct | 829 | |
| Db | 721 | GGAGCTGGTGAAGGTGGCCATTTCCCTCAGCATTTGACCGTGTGAGTTTACCTTCGTGA | 780 | |
| QY | 830 | tcctgtcaatcaaggttgacaacactcagctctacacttgggggccaagttgttggactaca | 889 | |
| Db | 781 | TCTGCCCATCAAGAGGTGACACCAATTCAGCTCTACCTGGGGGCCAAGTGTGTGGACTACA | 840 | |
| QY | 890 | gggaaagtgtgacaaagtgttccataactctgcagcttccctgtacaatgtcccaacctgga | 949 | |
| Db | 841 | GGGAAAGGTGACCAAGTGTGTAATTACTGTGCAAGCTTCCCTGCATCAATGCCCACCTGGA | 900 | |
| QY | 950 | caacatccgltcagactcactggaatcaggaagctgtgtgaagctgcaatgtgcctctct | 1009 | |
| Db | 901 | CAACATCCCGTTAGCCTCATCGTAGTCAAGAGTGTGAAGCTGTGCAATGCTGCTGT | 960 | |
| QY | 1010 | gctcttcacgaagaatctatgttccctgtttggaactctgtgtctctgtgaagtgtgccatcg | 1069 | |
| Db | 961 | GCTTCTTCAGAAATATTCATAGTCTCGTTGGACTCTGTGCTTCTCGAGAGTGGCCATCG | 1020 | |
| QY | 1070 | gcttgaatcacaagcatcgggtgtgatacaatgaagaagctgtcagataagctgtgatactacca | 1129 | |

| | | | |
|------------|--|---|------|
| Db | 1021 | GCTGAAGTCACGACATCGGGCGTGTATCAATGAAAGGCTGCATATAAGCTGGATCTACC | 1080 |
| Qy | 1130 | gatgttgaagttccttaacttcaggacatcccgagttttttaagaacaaggcatgc | 1189 |
| Db | 1081 | GATGTGAAGATCCTTAACCTCAGGACTCCCGAATTTTTTTTAAACACAAGGCCATTG | 1140 |
| Qy | 1190 | gattggccaactatgctgtctcygaagtgcttcccctcaagtgaagccccctggctt | 1249 |
| Db | 1141 | GGTGCCCAACTGATTCGTCGTGGAAGTGTTCCCTCCAGTCAACCCCTCCGCCCTT | 1200 |
| Qy | 1250 | caccctggcatcgaaagccagctctcgagctcagttttaaccaaaagtgaaccaat | 1309 |
| Db | 1201 | CACCTTGGGATCCAAACCCAGCTCGGAAGCTCATGTTTAAACCAACAAAGTGAC | 1260 |
| Qy | 1310 | actcaacttaataaacatcagctctgtatcgatccagctgaatgaaccttggattg | 1369 |
| Db | 1261 | ACTCAACTTGAATTAACATCACTGACTCTGTGATCGATCCAGCTATGAAATCTGG | 1320 |
| Qy | 1370 | gttccaaactgaatgttctgaaaaaacatcacatcaatgaatcatccaatccatgc | 1429 |
| Db | 1321 | GTTCACACCTGATGTGTTGAAAAAACATCACTAGATATATCACACTCATCTGCG | 1380 |
| Qy | 1430 | gaaccagaatgagaaataataagatctctgggtgccagtgatctgttgaagccttg | 1489 |
| Db | 1381 | GAACCGAAATGGCAAAATTAAGATCTGGGGTCCCAAGTGTATGGTGAAGGCTT | 1440 |
| Qy | 1490 | cgaagcagctgagcttcctcaactgaccaaagatgctgcctgtcttaccaccaac | 1549 |
| Db | 1441 | CGAGCGAGCTGAGATCCCATGACCAAAGATGCCCTTGTCTTACTTCCAGCTCT | 1500 |
| Qy | 1550 | gaaccagctctcctctgtcttccagtgaaacttgatgagcgcacatcaggaagctg | 1609 |
| Db | 1501 | GAACCCAGCTCCCTCCGTCCAGTGAAGAAGCTTGATGTCAGCAGCATCAAGGCTG | 1560 |
| Qy | 1610 | gtcccagctggagatgttgtgagctctatagaccatccccctctgcaatcaataac | 1669 |
| Db | 1561 | GTCCACAGCTGGAGATTTGGGTGAGACTCTATAGACCATCCCTCTGCAATCAAT | 1620 |
| Qy | 1670 | aactgacctgtg 1680 | |
| Db | 1621 | ACTTGCTGTG 1631 | |
| RESULT | 4 | | |
| LOCUS | BC008429 | 1684 bp mRNA linear PRI 12-JUL-2001 | |
| DEFINITION | Homo sapiens, Similar to DNA segment, Chr 2, Massachusetts Institute of Technology 19, clone MGC:14597 IMAGE:4291561, mRNA, complete cds. | | |
| ACCESSION | BC008429 | | |
| VERSION | BC008429.1 GI:14250057 | | |
| KEYWORDS | MGC. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 1684) | | |
| TITLE | Strausberg,R. | | |
| JOURNAL | Direct Submission Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA | | |
| REMARK | NIH-MGC Project URL: http://mgc.nci.nih.gov | | |
| COMMENT | Contact: MGC help desk Email: cgabs@email.nih.gov Tissue Procurement: CLONTECH cDNA Library Preparation: CLONTECH Laboratories, Inc. Genome Sequencing by: The I.M.A.G.E. Consortium (LNL) Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www.shgc.stanford.edu | | |

Contact: (Dickson, Mark) med@axil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at: <http://image.llnl.gov>
Series: IRM Plate: 21 Row: 3 Column: 12
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES

source

1. .1684

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:14597 IMAGE:4291561"

/tissue_type="Skeletal Muscle"

/clone_11b="NH_MGC_81"

/lab_host="DH10B"

/note="Vector: pDNR-LIB"

77. .1531

/codon_start=1

/product="Similar to DNA segment, Chr 2, Massachusetts

Institute of Technology 19"

/protein_id="AAH08429.1"

/db_xref="GI:14250058"

/translation="MAGPWTFLLGSLAATLLOATLSPTAVLIIGPKYIKELNOEL

KDHNTSTIOLPLLSAMREKRGISVYVLTVMKXITANILLOVRRPS

ANDDELKIPIDMYAGRTPLVKTIVEFHMTTEQATIRMTSASGPRILSDCAT

SHGSLRIQLHLKLSFLVNALAKQVNLVPSLPNKNQCVIAESFGMTADLLQL

VKPISLSIDRLFDILYPAIKQDITQIYLGKILDSOGKYKWNNSAASLTMPILD

NIPSLIVSODVKAANAVALSPSEFMYLDLSVESARLSKISGLINEKADKGS

TQIVKILQDPEFPIIDOGHAKVAOLIVLEVPSSEALRPLFTLIESSAEQFTKG

DOLILNINNISDRIOQMNSGIMGFQPVYKLIITETIHSILIPNONGKLRGVSVL

VKALFEAESSLSDALVLTPLTSMKPTSPVSQ"

BASE COUNT 419 a 493 c 414 g 358 t

ORIGIN

Query Match

Best Local Similarity 99.7%; Score 1615.8; DB 9; Length 1684;

Matches 1628; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

49 gagtgaagcggcgaggaactcagcgtgccagctgtgacatcctgtgaccttgcct 108
111:|||||
2 gagagagggggcgaggaactcagcgtgccagctgtgacatcctgtgaccttgcct 61
109 ctgacacttgggaagaatgagccggccgtggaacttacccttctgtgtgtgtgaa 168
111:|||||
62 ctgacacttgggaagaatgagccggccgtggaacttacccttctgtgtgtgtgaa 121
169 ggaacattgatacaagccaccctcagctccacatgacatcttctctgagccaaagc 228
111:|||||
122 gccaccttgaaccacccacctcagctccacatgacatcttctctgagccaaagc 181
229 atcaaaagaagctgacacagagctgaagagacacacagccacagcatcctgcagac 288
111:|||||
182 atcaaaagaagctgacacagagctgaagagacacacagccacagcatcctgcagac 241
289 ctgcgcgtctcagtgatgcatacgaggaagaagccagcgagag-atcctgtgtgtgagc 347
111:|||||
242 ctgcgcgtctcagtgatgcatacgaggaagaagccagcgagag-atcctgtgtgtgagc 301
348 ctggtgaagaacgctctgagacacatcattgtgtgaagatatacaagcatcaatcctc 407
111:|||||
302 ctggtgaagaacgctctgagacacatcattgtgtgaagatatacaagcatcaatcctc 361
408 cagctgcaagtgtaagccttcgagcaatgacacagagctgtgataagatcccccagac 467
111:|||||
362 cagctgcaagtgtaagccttcgagcaatgacacagagctgtgataagatcccccagac 421
468 atgtgtgtgtgtgtatcaaacagccctctgtcaagaacatcgtgtgagttccaatgacact 527
111:|||||
422 atgtgtgtgtgtgtatcaaacagccctctgtcaagaacatcgtgtgagttccaatgacact 481

528 gagggcccaagccacatccgcatgagacacacagtgagccccaccgctgtgtctc 587
111:|||||
482 gagggcccaagccacatccgcatgagacacacagtgagccccaccgctgtgtctc 541
588 agtgaactgtgcccacagccatgagagcctgtgagatcccaactgtgtgataagctctc 647
111:|||||
542 agtgaactgtgcccacagccatgagagcctgtgagatcccaactgtgtgataagctctc 601
648 ctgtgtaagccttgaagcagtgatcaatgacatccctcctagtgacatccctgtgacatc 707
111:|||||
602 ctgtgtaagccttgaagcagtgatcaatgacatccctcctagtgacatccctgtgacatc 661
708 gttaaaaaacagctgtgtcccgatgacagagcctcctcctcaatgtgatatgagaagcct 767
111:|||||
662 gttaaaaaacagctgtgtcccgatgacagagcctcctcctcaatgtgatatgagaagcct 721
768 ctgcaagctgtgtaagtgcccatctccctcaagcatgacccctgtgagttgacctctg 827
111:|||||
722 ctgcaagctgtgtaagtgcccatctccctcaagcatgacccctgtgagttgacctctg 781
828 tatcctgtgcatcaagtgatgacacatcagcttaccctgtgagggccaagtgtgtgactca 887
111:|||||
782 tatcctgtgcatcaagtgatgacacatcagcttaccctgtgagggccaagtgtgtgactca 841
888 cagggaaaggtgacaaagtgtgtcctcaatcctcctgagcctcctgagcaagccacccctg 947
111:|||||
842 cagggaaaggtgacaaagtgtgtcctcaatcctcctgagcctcctgagcaagccacccctg 901
948 gaaacaatcccgctcagcctcagctcagtgagtcaggaacgtgtgaaagcctgacgtgtcct 1007
111:|||||
902 gaaacaatcccgctcagcctcagctcagtgagtcaggaacgtgtgaaagcctgacgtgtcct 961
1008 gtgtctctcagaaagaatcagtgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1067
111:|||||
962 gtgtctctcagaaagaatcagtgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1021
1068 cagctgaagtgacagcctgagcgtgacatgaaagcctgtgaaagcctgtgagatgac 1127
111:|||||
1022 cagctgaagtgacagcctgagcgtgacatgaaagcctgtgaaagcctgtgagatgac 1081
1128 cagatcgtgaaagatcccaactcaggaacatcccgagttttttaagaacaaagccatgac 1187
111:|||||
1082 cagatcgtgaaagatcccaactcaggaacatcccgagttttttaagaacaaagccatgac 1141
1188 aagtgagcccaactgacgt 1247
111:|||||
1142 aagtgagcccaactgacgt 1201
1248 ttcaacctgtgagcctgaaagcagctcggaagcctcagtttttaacaaagtgacaaatt 1307
111:|||||
1202 ttcaacctgtgagcctgaaagcagctcggaagcctcagtttttaacaaagtgacaaatt 1261
1308 atactcaacttgataatacatcagctgtgacatcaggaatcaggaatcaggaatcaggaatc 1367
111:|||||
1262 atactcaacttgataatacatcagctgtgacatcaggaatcaggaatcaggaatcaggaatc 1321
1368 tggttccaactgt 1427
111:|||||
1322 tggttccaactgt 1381
1428 ccgaacccaagtgacaaatgaagatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1487
111:|||||
1382 ccgaacccaagtgacaaatgaagatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1441
1488 ttgcaagcagctgagctcctcaactgacacaggaatgacatcctgtgtgtgtgtgtgtgtgtgt 1547
111:|||||
1442 ttgcaagcagctgagctcctcaactgacacaggaatgacatcctgtgtgtgtgtgtgtgtgtgt 1501
1548 tggaaacccaagctcctcctgt 1607
111:|||||
1502 tggaaacccaagctcctcctgt 1561

QY 1608 ggtccacagctggagatagtggtgagctctataagaccatccctctctgcaatcaataa 1667
|||||
Db 1562 GGGGCCAGCGTGGGAGATGGGTGAGCTCTATAGACCATCCCTCTCTCATCAATTA 1621
|||||
QY 1668 aactgctctgtg 1680
|||||
Db 1622 ACACCTGCTGTG 1634
|||||

RESULT 5
MMU46068 1629 bp mRNA linear ROD 10-AUG-2000
LOCUS Mus musculus von Ebner minor salivary gland protein mRNA, complete
DEFINITION
ACCESSION U46068
KEYWORDS U46068.2 GI:9789706
SOURCE house mouse strain-Swiss-Webster.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1629)
Snead, M.L., Villanueva, J., Paine, M.L., Lei, Y.P., Zhu, D.H.,
Lusis, J., Xia, Y.-R. and Yang, J.-N.
TITLE Submitted (17-JAN-1996) Malcolm L. Snead, Center for Cranio
Molecular Biology, University of Southern California, 2250 Alcazar
St, Los Angeles, CA 90033, USA
COMMENT On Aug 10, 2000 this sequence version replaced gi:1184789.
FEATURES
Location/Qualifiers
1..1629
/organism="Mus musculus"
/strain="Swiss-Webster"
/db_xref="taxon:10090"
/chromosome="2"
/map="2; DZM119n and DZM125n"
58..1482
/codon_start=1
/product="von Ebner minor salivary gland protein"
/protein_id="AA87581.2"
/db_xref="GI:9789707"
/translation="MAGPWITLGLGATLVQAVNYPPAVNLGPEVYLOKHLQAL
KDHDAIPILOELPLRAMODSGSIPIDSEFVHTVLYKITIMKVTSANILQDVOST
YNOELVRIPLDMWAGLNPILKIVEFQSTVOLAIRERKSGSPAHNLNDSGSN
ESTRLSLHLKLSFVNSLAKNYMNLVPAIPQIVNNHLCPMIQAFDMYEDFLRLT
TAPALSPGLERGLSPALQDSNITLNLAKLDSQARTNMNFNSATSLMETPFDG
APPSLYRQDLVNAIVTTLVPEKEELVTLFEVPIPDVROLDIKEINAEANKLQPT
QMLKTFHSTPHVILNEGSARAQSVLVLEFPNTIDVREFPSIGISAEQAFPEEDN
RLMLNFMNNSVIERIKIMISDIKLEFDEVLKDTLTKILLEYTLIPNENKLRGTIPMSMS
KALGYERAMSVSKALKITLPASS"

BASE COUNT 411 a 497 c 382 g 339 t
ORIGIN

Query Match 49.0%; Score 822.8; DB 10; Length 1629;
Best Local Similarity 71.2%; Pred. No. 8.5e-185;
Matches 1129; Conservative 0; Mismatches 452; Indels 5; Gaps 3;

QY 96 cactgctgcctcgaacacttgagaaatgagccgagctggaaacttaacctctctg 155
|||||
Db 30 CACCTACTTCTACGCGACCCAGAGAGATGGCCGCCGTGATTATCACTCCTCTG 89
|||||

QY 156 tggatttgctgcaagcaacttgatcaagcaccctcagtcacacagagttctactct 215
|||||
Db 90 TGGTTTGGTGGAGCCACTGTGCAAGCTACTCTTCCCTGGCAGTGTCAACT 149
|||||

QY 216 cggcccaaaagtatcaagaagaactgacacagagctgaagagacacaacgcaacag 275
|||||
Db 150 TGGCCAGGAAGTATCAGAAACACCTGACCCAGGAGCACTGAAGAGACATGATGCCATCC 209
|||||

QY 276 catccccagcagcagctgcgctgctgctgctgctgaggaagaacgagccgagatccct 335
|||||
Db 210 CACTCTCCAGGAGTTCCTCACTGCTCAGAGCCATGCAAGATAGTCTG--GCAGTATCCCC 267
|||||

QY 336 gtctgagcagcctgtgtgaacaacgcgtctgaagcacatcatctgtcgtgaagtatcaca 395
|||||
Db 268 ATACTGACAGCTGTGTGCACACCTTCTGAAATATCATATGATGATGAGTCACTCT 327
|||||

QY 396 gctaacatctccagctgcaagtgtaagccctcggccaatgacccaagagctgtagtaag 455
|||||
Db 328 GCTAACATCTCCAGCTGGATGTGACGCTTCGACCTATACCCAGGAGCTGTGTGAGA 387
|||||

QY 456 atccccctggaactggtgctgagatctaaacagccccctggttcaagacactcgtgagatc 515
|||||
Db 388 ATCCCGCTGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 447
|||||

QY 516 cacatgacagctgagcccaagccacacatccgcatgagacacagctgtaagtgccccacc 575
|||||
Db 448 CAATGAGACACCGAGGTCCACGCCCTCATCCGGTGGAGAGAGACCAAGAGCGGCCGCC 507
|||||

QY 576 cgctgtctcctcagtgactgtgcccacgaagcgaagagagctgcaatcgaatgctgcat 635
|||||
Db 508 CACTGGAATCTCAGCGACTGCTCAGCAATGAGAGACCCCTGCGCTCAGCTGCTTCA 567
|||||

QY 636 aagctctcctcctggtggaagcgccttaagtaagcagtgtaagctcctcagtgccatcc 695
|||||
Db 568 AAGCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 627
|||||

QY 696 ctgcccacatagtgaaacacacagctgtgtcccgtagtcgaagctcctcctcaatgagcatg 755
|||||
Db 628 CTGCCCCAATATGTAAGAAACACCTATGCCCTATGATGATGATGATGATGATGATG 687
|||||

QY 756 tatgcaagactcctgcaagctgtggaagtgagcctcctcctcctcctcctcctcctgag 815
|||||
Db 688 TAGGAAGACTTCTGAGACTGACACACACACCATTTGCCCTGAGCCCTTGGAG 747
|||||

QY 816 ttgacactctgtatctcctgcaacaggtggaacacacacacacacacacacacacacac 875
|||||
Db 748 TTTGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 807
|||||

QY 876 ttgtgagctcaagaggaagtggaagtggaagtggaagtggaagtggaagtggaagtggaag 935
|||||
Db 808 CTGCTGACTCACAGGCGAGGAGTAACCACTGTTCAACACTGCGCACTCTCTGATG 867
|||||

QY 936 atgcccacctggaac 995
|||||
Db 868 GAGACACACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 927
|||||

QY 996 gcaagtgctgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1055
|||||
Db 928 ATTGTGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 987
|||||

QY 1056 gtagatgcccac 1115
|||||
Db 988 GATGTGGCCCGCCAGTTCACAGATGACATCAAGAAATCAATGACAGAGGAGCCAGCAAG 1047
|||||

QY 1116 ctggagatctac 1175
|||||
Db 1048 CTGGGGCCACCCACATGTTACAGATCTTACACCAACACACCCCCACATGTGCTGAT 1107
|||||

QY 1176 caagac 1235
|||||
Db 1108 GAGGCACTGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1167
|||||

QY 1236 ctgcgccccttgctac 1295
|||||
Db 1168 GTCCGGCCCTTCT 1227
|||||

QY 1296 ggtgacacacttaactcaacttgtaataacatcaagctcgtgacagctgagatgagac 1355
|||||
Db 1228 GACAACCGGCTTATGCTCAACTTCAATATATGCTGATGAGGAGCATCAAGCTGATGATC 1287
|||||

QY 1356 tctggagatgctgctgcaac 1415
|||||
Db 1288 TCAATATATCAACTATTTGATCTCTGAGAGCTGTAAGAGACACCCCTGACCAAGATCTTGA 1347
|||||

| QY | 1416 | tcacacctgctgcgcgaacccagaatggaatlaagaatctgggtgccagttcatggtg | 1475 |
|---------------------------|---|---|--------------------|
| Db | 1348 | TACACACTGCTGCCCAATGAAATGGCAACTGGAATCGAATCCCATGCAATGCC | 1407 |
| QY | 1476 | aagccctggagatcgaagcagctgagtcctccactgagcaagagatgacctgtgtact | 1535 |
| Db | 1408 | AAGCCCTTGGGATACGAGAAAGCCATGTGTCTGTGACCAAGGGTGCCCTCAAGCTTCACT | 1467 |
| QY | 1536 | ccagcctcctctgttggaaaccagctctcc-tgctcccaagtgaagacttgatgcaagc | 1594 |
| Db | 1468 | CCAGCCTCTCTCTGAAGAACCTTGCGTCCCCCGCCCTCTGTGATGAAGACCTGGACATGACCA | 1527 |
| QY | 1555 | atcagggaaagctggtgtccagctcgtgagatgtggtgtgtgactctatagaccatccctc | 1654 |
| Db | 1528 | GCCAGCCCCCGCCGAGTCCCGACGAGGAGTGTGGAAAGCCTGTGTAGAC--TACCCCC | 1585 |
| QY | 1655 | ctgcaatcaataaactgctgctgtg | 1680 |
| Db | 1586 | CTGCAGTTAATTAACACTTGGCCCGTG | 1611 |
| RESULT | 6 | | |
| AX301908 | | | |
| LOCUS | AX301908 | 1035 bp | DNA |
| DEFINITION | Sequence 15 from Patent WO0174851. | | Linear |
| VERSION | AX301908 | | |
| KEYWORDS | AX301908.1 GI:17382958 | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eulheria; Primates; Catarrhini; Homnidae; Homo. | | |
| | 1 (sites) | | |
| | Mjunder,K., Spaderna,S.K., Taupier,R.D., Padigaru,M., | | |
| | Burgess,C.E., Shinkets,R.A., Spytek,K.A., Liu,X., Patturajan,M. and | | |
| | Gusev,V.Y. | | |
| TITLE | Novel proteins and nucleic acids encoding same | | |
| JOURNAL | Patent: WO 0174851-A 15 11-Oct-2001; | | |
| FEATURES | Curagen Corporation (US) | | |
| source | Location/Qualifiers | | |
| | 1..1035 | | |
| CDS | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | 79..1035 | | |
| | /note="unnamed protein product" | | |
| | /codon_start=1 | | |
| | /protein_id="CA013090.1" | | |
| | /translation="MAGPPTFTLLCGLAATLIDATISPRNAVILIGKVIKEKITQEL | | |
| | KDHNAISILLOOLPLLSAMREKPAIGIVPLSLVNTVILKHIIMLVITANILQIQKPS | | |
| | ANDQELVKKIPLDMVAFNPPLVKTIVIEFHTTEAQTIRMDTSASPTRIYLSDCATM | | |
| | SHSIRIQLHLKSLFVINALAKQVMAIVLSPILVKNQICLVTEASFNQYADLLDLO | | |
| | VKSGASLSPTFSFTTFLASRPCKVTKWPNNSASLIMPTIDNIPFLIVISQDYVKKAA | | |
| | VAAVLSPEERMWILDSVYNLSTRGRICPPRPRIHNPFLNTGCP" | | |
| BASE COUNT | 230 a 333 c 261 g 211 t | | |
| ORIGIN | | | |
| Query Match | 48.0%: | Score 806: | DB 6: Length 1035; |
| Best Local Similarity | 91.5%: | Pred. No. 8.2e-181; | |
| Matches 906; Conservative | 0; | Mismatches 35; | Indels 49; Gaps 3; |
| QY | 63 | gaggactccaagcgtgcccgaagtctggaactctgactgctgagcctctgacacctggaa | 122 |
| Db | 18 | GAGGACTTCACAGCGGCCAGGCTGCGATCCTGCACCTTACGTGCTGACACCTGGGA | 77 |
| QY | 123 | gatggcggcccgctggagccttcaccctctctgtgtgttctggtgagcagccattatcca | 182 |
| Db | 78 | GATGGCGCGCCGCGAGACCTTCACCTCTCTGTGTGTTGCTGGCAGCACCATTATCA | 137 |
| QY | 183 | agccacctgagtcaccatgcaagtctcatctctggtgcccacaaagtcatcaagaagaagct | 242 |
| Db | 138 | AGCCACCTTCAGTCCCATGTGAGTTCATCTCTCGGCCCAAAAGTCATCAAGAAAGCT | 197 |

| | | | |
|----|------|--|------|
| OY | 243 | gacacaaggagctgtaaagaaccacaacycaaccagcatcctgtcaagcagctgtgcgctgtcag | 302 |
| Db | 198 | GACACAGGAGCTGAAGGAGCACAAACGCCACCAGATCCTGGCAGCATGCCTGTCTCAG | 257 |
| OY | 303 | tggccatgtgggaagaagccaagcccggaggt-attcccttgttggcgacccctgtgtgaacacct | 361 |
| Db | 258 | TGCCATTGGCGGAAAAAGCCAGCCGGAGGAGATCCCTTGTGCTGGCCACCTGTGTAACACCGT | 317 |
| OY | 362 | ccttgagacacatcatctgtgttaagtatcacagactaacatccctcccaagctgtcagttgaa | 421 |
| Db | 318 | CCTGAAGCAATCATCTGGCTTAAGTGTCATCACACTAACATCCTCCAGCTGCAGGTGAA | 377 |
| OY | 422 | gcccttcggccaattgaccagagagctgtcgttcaagaatccccctgtgscatgtgtgcgtgalt | 481 |
| Db | 378 | GCCCTTCGGCCCAATGACCCAGAGACTCTGTAGTMAAGTCCCCTGGACATAGTGGCTGAGATT | 437 |
| OY | 482 | caaacagccccctgtgtccaagaccaatgtgtgaagtgtccaatgtgaagactgtggagcccaagcac | 541 |
| Db | 438 | CAACACGCCCTGTGTCAAAGACCATGTGGTAGTTCCACATGACGACTGTAGGGCCAAAGCAC | 497 |
| OY | 542 | catccgcatalygacacacagltgcaagtgtgcgcccaaccccgcctgtgtcctcaagtgtcgtgcac | 601 |
| Db | 498 | CATCCGCATGTGACACAGAGCAAGTGGCCCCACCCCGCGTGGCHCNAGAGACTGTGGCAC | 557 |
| OY | 602 | cagccatatggagcctctgcgtgatcccaactctgtcatagaetctccctctcgtgtgaagcctt | 661 |
| Db | 558 | CAGCATATGGGACCCCTGGCGATCCAACTCGTGTATMAAGCTCCTCTTCGTGTAAAGCCCTT | 617 |
| OY | 662 | agctaagcaggtgtcatgaacctcttagtgcatacctctgcacacacacacacacacacacac | 721 |
| Db | 618 | AGCTAAGCAGGTGATGAACTCTCTAGTGTGCATCCCTCGCCAACTTAGTGAATAAACACACT | 677 |
| OY | 722 | gtgtcccgatgatcgaggtcttccatcaatlygaatgtatgcagaacccctcgtcagctgtgtgaa | 781 |
| Db | 678 | GTGTCCCGGTGATGAGAGGCTTCCTCAATGGCATGTATGACAGACCTCCAGACTGTGTAA | 737 |
| OY | 782 | ggtgtgccaattccccacagcatgtgacagctgtgagtttgagctctctgttatccgtgcataaa | 841 |
| Db | 738 | GG-----GTAGGTGCTGTGCTCTCTCTCCACACTTTTTCT----- | 772 |
| OY | 842 | gggtgacacacatcagctctacactgtgggggccaagttgttgagactacagagggaaaagttgac | 901 |
| Db | 773 | -----TTACTACGGAGCTGGCGCTTCAGACCCCGGAAAAGGTGAC | 809 |
| OY | 902 | caagtggttcaataactctgtcagcttccctgtgcaatgtgccacccctggaaaaaattccggt | 961 |
| Db | 810 | CAAGTGTTTCATTAACCTCTGCAGCTTCCCTGCATCAATGCCACCCCGGACAATATCCGTT | 869 |
| OY | 962 | cagctcatctgtgagtgcaagagacgtgtgtlaaaactcagdtgtgtgtgtgtcctctccaga | 1021 |
| Db | 870 | CAGCTCATCTGTGAGTCAAGAGAGCTGTGTAAAGCTGCAGTGGCTGCTGTCTCTCTCCAGA | 929 |
| OY | 1022 | agaattcatgtgtcctgttggactctgtgtct | 1051 |
| Db | 930 | AGAATTCAATGGTCTGTGGACTCTGTGTGT | 959 |

RESULT 7

LOCUS AX301906 963 bp DNA linear PAT 30-NOV-2002

DEFINITION Sequence 13 from Patent WO0174851.

ACCESSION AX301906

VERSION AX301906.1 GI:17382956

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (sites)

Majumder,K., Spaderna,S.K., Taupier,R.J., Padigar,M.,

Bujages,C.E., Shimkels,R.A., Spytek,K.A., Liu,X., Patturajan,M. and

Gusev,V.Y.

[illegible]

| QY | 837 | atcaagggtgcacacattcagctcactacctgaggggccaagttgttggactacacagaaggaag | 896 |
|------------|--|---|------|
| Db | 701 | -----TTACTACGAGCACTGCGCTCCACAGACCCGCAAGG | 732 |
| QY | 897 | gtaccacaagtgttcataactcactgcagctccctctacacatgcccacccctggaacaac | 956 |
| Db | 733 | gtgacacaagtgttcataactcactgcagctccctctacacatgcccacccctggaacaac | 792 |
| QY | 957 | ccgttcacacccctcactgcagctcaggaacgtggtgaaagctgcagctgctgtgctct | 1016 |
| Db | 793 | ccgttcacacccctcactgcagctcaggaacgtggtgaaagctgcagctgctgtgctct | 852 |
| QY | 1017 | ccagaagaatcctcagctccctgttgagactctgtgct | 1051 |
| Db | 853 | CCAGAAGAATCCTCAGCTCCTGTGAGACTCTGTGCT | 887 |
| RESULT | 8 | | |
| LOCUS | AL355392 | | |
| DEFINITION | Human DNA sequence from clone RPS-1187/4 on chromosome 20q11.1-11.23 Contains ESTs, STSS, GSS and two CpG islands. | | |
| ACCESSION | AL355392 | | |
| VERSION | AL355392.7 | | |
| KEYWORDS | HTG; CGI-05; CpG island; RYF3; SNT1; SNTA1; syntrophin. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 159440) | | |
| TITLE | Bird.C. | | |
| JOURNAL | Direct Submission | | |
| COMMENT | Submitted (06-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk On May 14, 2001 this sequence version replaced gi:5982468 gi:5213549. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WormRep; Information on the WormRep database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormrep This sequence was generated from part of bacterial clone configs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20 This sequence is the entire insert of clone RPS-1187/4 The true left end of clone RPS-1187/4 is at 62286 in this sequence. The true right end of clone RPS-1187/4 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RPS-1187/4 is from the library RPS-5 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pCYPAC2. | | |

| FEATURES | | Location/Qualifiers |
|---------------|--|--|
| source | | 1..159440 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="20" /map="q11.1-11.23" /clone="RP5-1187J4" /clone_1fb="RPCI-5" 1535..1848 /note="157 copies 2 mer ga 61% conserved" |
| repeat_region | | 1757..1824 /note="Random repeat. weak data" join(3408..3471,5782..5867,7116..7295,8242..8295, 8799..8957,9841..9908,10729..10774,11885..11948, 12794..12870,15591..15761) /gene="dJ1187J4.1" |
| misc_feature | | /note="match: ESTs: Em:BE933124 Em:BE061388" /product="dJ1187J4.1.1 (novel protein similar to mouse von Ebner salivary gland protein, isoform 1.)" |
| gene | | /evidence="not_experimental" 3408..15761 /gene="dJ1187J4.1" |
| CDS | | join(<3408..3471,5782..5867,7116..7295,8242..8295, 8799..8957,9841..9908,10729..10774,11885..11948, 12794..12870,15591..15650) /gene="dJ1187J4.1" /note="Continues in Em:AL121901 as bA49610.6" |
| | | /codon_start=1 /evidence="not_experimental" |
| misc_feature | | /product="dJ1187J4.1.1 (novel protein similar to mouse von Ebner salivary gland protein, isoform 1.)" |
| protein_id | | /protein_id="CAC34050.1" |
| db_xref | | /db_xref="GI:13274680" |
| translation | | /translation="LCPTVEASFNGMVRADLIQLVKVPISLIDRLPDLTPAIKQDT IQTLGATLDSQGVKTFENNSASLTPLDNPESLIYSQVKAAVAIVSPER FMVLDSVLPESARLKSSIGLINERKADKLGSTQIVYLIDPPEPFIDGHAQVQ LIVLEVPSSSEALRPLETLGIETASAEQYTGQQLILNNISSDRLOLNMSGWF OPDVAKNIITEIHISILPNQNGKLRSGVPLVKALGFEEAESLTKDALVLPASL WKSSPVSO" |
| misc_feature | | 3817..3915 /gene="dJ1187J4.1" |
| misc_feature | | /note="match: STS: Em:G10242" complement(join(3917..4006,31119..31326)) /note="match: STS: Em:U75536" |
| misc_feature | | 3917..4143 /gene="dJ1187J4.1" |
| repeat_region | | /note="match: STS: Em:G10242" 4662..4691 /note="15 copies 2 mer tt 86% conserved" |
| misc_feature | | join(5758..5867,7116..7295,8242..8295,8799..8957, 9841..9908,10729..10774,11885..11948,12794..12870, 15591..15761) /gene="dJ1187J4.1" |
| misc_feature | | /note="match: CDNAS: Em:U46068 match: ESTs: Em:A1924632 Em:AM363590 Em:AM384404 Em:AM449208 Em:AM167650 Em:AM381927 Em:AM167610 Em:AM577643 Em:AM176323 Em:A1024451 Em:DA5555 Em:AM176329 Em:AM176327 Em:AM176325 Em:A1909965 Em:A1909964 Em:AA428580 Em:AM384405" |
| misc_feature | | /product="dJ1187J4.1.2 (novel protein similar to mouse von Ebner salivary gland protein, isoform 2.)" |
| misc_feature | | /evidence="not_experimental" complement(6761..7253) /note="match: GSS: Em:AQ149447" |
| CDS | | join(7182..7295,8242..8295,8799..8957,9841..9908, 10729..10774,11885..11948,12794..12870,15591..15650) /gene="dJ1187J4.1" |
| | | /codon_start=1 /evidence="not_experimental" |
| misc_feature | | /product="dJ1187J4.1.2 (novel protein similar to mouse von Ebner salivary gland protein, isoform 2.)" |
| protein_id | | /protein_id="CAC15880.1" |
| db_xref | | /db_xref="GI:11137741" |
| | | /translation="MPTLDNIPFSLIVSQDVKAAVAIVSPFEFNLDSVLPESAR LKSSIGLINERKADKLGSTQIVKILIDPPEPFIDGHAQVLAQVLIYEVPSSSEAR PLPTGIETASAEQYTKGQDLILNNISSDRLOLNMSGWFOPVAKNIITEIHS SILPNQNGKLRSGVPVSLVALGFEAESLTKDALVLPASLWKSSPVSO" |
| repeat_region | | 10970..11009 /note="20 copies 2 mer ac 85% conserved" |
| misc_feature | | complement(11466..11766) /note="match: GSS: Em:A0060606" |
| misc_feature | | complement(11631..11769) /note="match: GSS: Em:A0427662" |
| misc_feature | | 11766..12192 /gene="dJ1187J4.1" |
| misc_feature | | /note="match: GSS: Em:AQ403818" |
| misc_feature | | 11768..12108 /gene="dJ1187J4.1" |
| misc_feature | | /note="match: GSS: Em:AQ373301" |
| misc_feature | | 11775..12108 /gene="dJ1187J4.1" |
| misc_feature | | /note="match: GSS: Em:A0662246" |
| repeat_region | | 14322..14833 /note="256 copies 2 mer ca 68% conserved" |
| repeat_region | | 14836..14871 /note="18 copies 2 mer ac 88% conserved" |
| misc_feature | | complement(15031..15459) /note="match: GSS: Em:A0694930" |
| polyA_signal | | 15736..15741 /gene="dJ1187J4.1" |
| polyA_site | | 15754 /gene="dJ1187J4.1" |
| polyA_site | | 15761 /gene="dJ1187J4.1" |
| misc_feature | | complement(17559..18000) /note="match: GSS: Em:AQ349609" |
| misc_feature | | complement(17617..18004) /note="match: GSS: Em:A0350711" |
| misc_feature | | 18109..18565 /note="match: STS: Em:HS812D7T" |
| misc_feature | | complement(20661..21112) /note="match: GSS: Em:A0019396" |
| misc_feature | | complement(20698..21113) /note="match: GSS: Em:AQ178024" |
| misc_feature | | 21105..21680 /note="match: STS: Em:HS156A22T" |
| misc_feature | | complement(22577..22968) /note="match: GSS: Em:AQ508829" |
| misc_feature | | 22977..23573 /note="match: GSS: Em:AQ281488" |
| misc_feature | | complement(37780..38128) /note="match: STS: Em:HSPF13F11" |
| misc_feature | | complement(39112..39718) /note="match: GSS: Em:A0636647" |
| misc_feature | | complement(39266..39748) /note="match: GSS: Em:A0827010" |
| misc_feature | | 39747..40042 /note="match: GSS: Em:AQ282137" |
| misc_feature | | 39754..40132 /note="match: GSS: Em:AQ252720" |
| misc_feature | | 39763..40037 /note="match: GSS: Em:AQ029439" |
| misc_feature | | match: STS: Em:G50113" |
| misc_feature | | complement(39950..40165) /note="match: GSS: Em:AL207848" |
| misc_feature | | complement(39950..40334) /note="match: GSS: Em:AQ556676" |
| misc_feature | | 39961..40217 /note="match: GSS: Em:AQ143633" |
| misc_feature | | complement(39964..40169) /note="match: STS: Em:G02503" |
| misc_feature | | complement(39991..40409) /note="match: GSS: Em:AQ596761" |
| misc_feature | | complement(40002..40172) /note="match: GSS: Em:A185783" |
| misc_feature | | complement(40006..40317) |

/note="match: GSS: Em:AQ052600"

Query Match 10.8%; Score 181.4; DB 9; Length 159440;
 Best Local Similarity 99.5%; Pred No. 2.2e-32;
 Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 869 gcccaagttgttgagtcacagggaaagtgacccaagtgttcaataactctgcagcttc 928
 |||||||
 Db 7115 GGCCAAGTTGTTGACTCACAGGAAAGTGACCAAGTGTTCAATCACTTCGACACTTC 7174
 |||||||

QY 929 ccgtgacacagccaccccttgacacacatcccttgacccatcctgtgagtcagacgtgt 988
 |||||||
 Db 7175 CCTGACAATGCCACCTTGACAAATCCGTTTCACCCATCAGTCAAGACAGCTGT 7234
 |||||||

QY 989 gaaagctgacgtgctgtctgtctctccacaagaatcatgctgctgtgagctgt 1048
 |||||||
 Db 7235 GAAAGCTGACGTGCTGTCTCTCTCTCCACAATTCATGCTGCTGTGTGACTGTGT 7294
 |||||||

QY 1049 gct 1051
 |||

Db 7295 GGT 7297

RESULT 9
 HSBAA9G10 161593 bp DNA linear PRI 20-JUN-2001
 LOCUS
 DEFINITION Human DNA sequence from clone RP11-49G10 on chromosome 20 contains a gene similar to bovine salivary protein BSP30, the LOC51297 gene for LUNX protein, 5' end of a gene encoding a protein similar to murine von ebner minor salivary gland protein, a novel gene, a pseudogene similar to STAT-induced STAT inhibitor-2, a pseudogene ESTs, STSs, GSSs and a Cpg island, complete sequence.
 A1121901
 A1121901.20 GI:8249854
 HTG: BSP30; Cpg island; LOC51297; LUNX; RPL12; STAT-inhibitor.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 161593)
 Tracey, A.
 Direct Submission
 Submitted (14-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquere@sanger.ac.uk
 On Jun 5, 2000 this sequence version replaced gi:8017404.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

COMMENT

IMPORTANT: This sequence is not the entire insert of clone RP11-49G10. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
 The true left end of clone RP11-49G10 is at 1 in this sequence. The true right end of clone RP5-1187J4 is at 161494 in this sequence. The true right end of clone RP4-733023 is at 27823 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were

covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-49G10 is from the library RPI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

FEATURES

source

Location/Qualifiers

1..161593
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RP11-49G10"
 /clone_lib="RPI-11.1"
 /complement(1..107)
 /note="match: GSS: Em:AQ067627"
 6..706
 /note="match: GSS: Em:AQ051854"
 137..548
 /note="match: GSS: Em:AQ550113"
 530..603
 /note="37 copies 2 mer ca 66% conserved"
 533..686
 /note="2 copies 77 mer 83% conserved"
 613..678
 /note="33 copies 2 mer ca 75% conserved"
 3030..3174
 /note="11PB3 repeat: matches 5941..6084 of consensus"
 5140..5183
 /note="22 copies 2 mer tt 75% conserved"
 5184..5344
 /note="L1PA2 repeat: matches 5986..6146 of consensus"
 complement(6104..6463)
 /note="match: GSS: Em:AQ339300"
 6120..6239
 /note="2 copies 60 mer 85% conserved"
 complement(8476..9002)
 /note="match: GSS: Em:AQ537190"
 complement(14766..15061)
 /note="match: STS: Em:HS922114S"
 14772..14779
 /note="tandem repeat. Forced join. Approximately 500 bases missing according to restriction digest."
 16219..16387
 /note="match: GSS: Em:AL001671"
 16225..16387
 /note="match: GSS: Em:AL013146"
 18325..18826
 /note="match: GSS: Em:AQ58133"
 18328..18828
 /note="match: GSS: Em:AQ532957"
 18333..18770
 /note="match: GSS: Em:AQ765930"
 complement(20975..21245)
 /note="match: GSS: Em:A2102442"
 20999..21238
 /note="4 copies 60 mer 69% conserved"
 21031..21263
 /note="match: GSS: Em:A2055779 Em:A2084220"
 21144..21225
 /note="41 copies 2 mer ag 75% conserved"
 complement(21172..21428)
 /note="match: GSS: Em:AQ395327"
 complement(23473..23950)
 /note="match: GSS: Em:AQ278777"
 complement(23637..23948)
 /note="match: GSS: Em:AQ691469"
 23950..24355
 /note="match: GSS: Em:AQ152593"
 23955..24456
 /note="match: GSS: Em:AQ508832"
 26979..27422
 /note="match: GSS: Em:AQ636646"
 31882..32061

/gene-"B4A9G10.2 (similar to STAR-induced STAT
inhibitor-2)
match: proteins: Tr:095102 Tr:088582 Tr:014542 Tr:014509
Tr:062225 Tr:015097 Tr:035718 Tr:090X77 Tr:090X78
Tr:070512 Tr:035960 Tr:09PW70 Tr:09YR1"

Query Match 9.6%; Score 161; DB %; Length 161593;
Best Local Similarity 100.0%; Pred. No. 1,6e+27;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 78 cccaagttgcacatccctgcaacttgcgtccttcgaaccctggagaagtgcgcccgctg 137
|||
Db 153404 CCCAGGTCTGGCATCCTTCACATTGCCTGCCCTTGACACTCGGAAGAATGCGCCGCTG 153463

Oy 138 gacctaaccccctctctgttgattgtctggaggacaaccttgtaccaagcacccccagtic 197
|||||
Db 153464 GACCCTCACCCCTTCTCTGTGGTTGGTCGCAAGCACCTTGATCCAAGCCACCTTAGTCC 153523

Oy 198 cactgcacattcatcactccgsgccsaagatcataaaana 238
|||||
Db 153524 CACTGCATTCTTCATCCTCCGCGCCCANAAGTATCAAATAAA 153564

RESULT 10
LOCUS AC094188/c
DEFINITION Rattus norvegicus clone CH230-2K16, *** SEQUENCING IN PROGRESS ***,
49 unordered pieces.
ACCESSION AC094188
VERSION AC094188.2 GI:17940901
KEYWORDS HTGS_PHAASEL.
SOURCE Norway rat.
ORANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Musculi.

REFERENCE
AUTHORS 1 (bases 1 to 89536)
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,G.,
Alsbrooks,S.L., Amerlunge,H.C., Are,J.R., Banks,T., Barbatala,J.,
Beaton,J., Birmage,K., Blankenbury,K., Bonnin,D., Boucek,J.,
Bowles,S., Britave,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Cartier,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Devila,M.L., Davis,C.,
Davy-Carroll,L., Dedereich,D.A., Delaney,K.R., Delgado,O.,
Demn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flieger,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guervara,W., Gunnaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognues,M., Holloway,C.,
Hollias,B., Homsi,F., Howard,S., Huber,J., Huylk.S., Hume,j.,
Jackson,L.E., Jacobson,B., Jia,y., Johnson,R., Jolyvet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovari,C., Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J.Z., Licharge,O., Lieu,C., Liu,D., Liu,W.,
Louisged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mamhinney,E., Meledon,M.P., Meadow,M.,
Mei,G., Metzker,M., Miner.G., Miner,Z., Mitchell,T., Mohabhat,K.,
Moragan,M., Morris,S., Moser,M., Neal,D., Newsum,J., Newton,N.,
Nguyen,A., Nguyen.N., Nguyen,N., Nickerson,E., Nwonkenwo,S.,
Ogunu,M., Okuwonu,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,
Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savory,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Stinson,I., Sodergren,E., Sonnikke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabori,P., Tameris,a,A., Tamerisa,K.,
Thang,H., Tansey,J., Taylor,C., Taylor.T., Tellrod,B., Thomas,N.,
Thomas,S., Usmani.K., Vasquez,L., Vera,V., Villation,D., Vinson,R,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Wellington, S., Williams, G., Williamson, A., Mieczys, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission

Unpublished
2 (bases 1 to 89536)

Worley, R.C.

Direct Submission

Submitted (17-SEP-2001)

Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi.15636537.

COMMENT

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GAEB

Center clone name: CH230-2K16

Summary Statistics

Assembly program: Phrap; version 0.990329first call to

findphraplist

Consensus quality: 73144 bases at least Q40

Consensus quality: 80017 bases at least Q30

Consensus quality: 85368 bases at least Q20

Estimated insert size: 69357; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 49 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 4713: contig of 4713 bp in length

4714 4813: gap of unknown length

4814 7442: contig of 2629 bp in length

7443 7543: gap of unknown length

7543 8970: contig of 1428 bp in length

8971 9070: gap of unknown length

9071 11618: contig of 2548 bp in length

11619 11718: gap of unknown length

11719 14064: contig of 2346 bp in length

14065 14164: gap of unknown length

14165 16598: contig of 2434 bp in length

16599 16699: gap of unknown length

16699 18558: contig of 1860 bp in length

18559 18658: gap of unknown length

18659 21636: contig of 2978 bp in length

21637 21736: gap of unknown length

21737 23091: contig of 1355 bp in length

23092 23191: gap of unknown length

23192 24363: contig of 1172 bp in length

24364 24463: gap of unknown length

24464 26616: contig of 2153 bp in length

26617 26716: gap of unknown length

26717 29328: contig of 2612 bp in length

29329 29428: gap of unknown length

29429 30861: contig of 1433 bp in length

30862 30961: gap of unknown length

30962 32666: contig of 1705 bp in length

32667 32766: gap of unknown length

32767 34751: contig of 1985 bp in length

34752 34851: gap of unknown length

34852 36722: contig of 1871 bp in length

36723 36822: gap of unknown length

36823 37990: contig of 1168 bp in length

37991 38091: gap of unknown length

38092 39851: contig of 1761 bp in length

39852 39951: gap of unknown length

39952 41483: contig of 1531 bp in length

41484 41582: gap of unknown length

41583 42970: contig of 1388 bp in length

42971 43071: gap of unknown length

43072 44717: contig of 1647 bp in length

44718 44817: gap of unknown length

44818 46731: contig of 1914 bp in length

46732 46831: gap of unknown length

46832 48721: contig of 1890 bp in length

48722 48822: gap of unknown length

48823 50484: contig of 1663 bp in length

50485 50584: gap of unknown length

50585 52334: contig of 1749 bp in length

52334 52433: gap of unknown length

52433 53718: contig of 1285 bp in length

53719 53818: gap of unknown length

53819 55081: contig of 1263 bp in length

55082 55181: gap of unknown length

55182 56437: contig of 1256 bp in length

56438 56537: gap of unknown length

56538 58395: contig of 1858 bp in length

58396 58495: gap of unknown length

58496 60176: contig of 1681 bp in length

60177 60277: gap of unknown length

60278 62186: contig of 1910 bp in length

62187 62286: gap of unknown length

62287 64498: contig of 2212 bp in length

64499 64598: gap of unknown length

64599 65989: contig of 1391 bp in length

65990 66089: gap of unknown length

66090 67493: contig of 1404 bp in length

67494 67593: gap of unknown length

67594 69641: contig of 2048 bp in length

69642 69742: gap of unknown length

69743 70862: contig of 1121 bp in length

70863 70962: gap of unknown length

70963 72389: contig of 1427 bp in length

72390 72490: gap of unknown length

72491 74083: contig of 1593 bp in length

74084 74182: gap of unknown length

74183 75218: contig of 1036 bp in length

75219 75318: gap of unknown length

75319 76809: contig of 1491 bp in length

76810 76909: gap of unknown length

76910 78708: contig of 1799 bp in length

78709 78808: gap of unknown length

78809 79849: contig of 1041 bp in length

79850 79949: gap of unknown length

79950 81538: contig of 1589 bp in length

81539 81638: gap of unknown length

81639 82748: contig of 1110 bp in length

82749 82848: gap of unknown length

82849 84058: contig of 1210 bp in length

84059 84158: gap of unknown length

84159 85695: contig of 1537 bp in length

85696 85795: gap of unknown length

85796 86875: contig of 1080 bp in length

86876 88063: contig of 1088 bp in length

88064 88163: gap of unknown length

88164 89536: contig of 1373 bp in length.

FEATURES

source

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone="CH230-2K16"

21918 a 20067 c 20357 g 22306 t 4888 others

BASE COUNT

ORIGIN

| | Matches | 119; | Conservative | 0; | Mismatches | 64; | Indels | 0; | Gaps | 0; |
|----|---------|--|--------------|----|------------|-----|--------|----|------|-------|
| QY | 1081 | gcacgagcgtacaaatgaagaagcgtgcagataagcttgatctaccagatcgtgaaga | | | | | | | | 1140 |
| Db | 38807 | GGATGTGTGTGATGCTTTCCAGGCGCAACAAAGTTGGGGCCACCCAGATGTTGAGAA | | | | | | | | 38748 |
| QY | 1141 | tcttaactaagaacactcccgagtttttatagacaaagccatgcgaagtggcccac | | | | | | | | 1200 |
| Db | 38747 | TCTTCACTACAGCACCCCCCATATTTTGTAAATGAGGGCGGTGCCAAGCGACGCCANA | | | | | | | | 38688 |
| QY | 1201 | tgatcgtcgtggaagtgatttccctccagtgaaagccctccgcctttgttaccctgggca | | | | | | | | 1260 |
| Db | 38687 | GGTGTGTAAGTGAATGTTCCCCACACCAATACAGATGTCCGACCCCTTCTTCCTGGCA | | | | | | | | 38628 |
| QY | 1261 | tccg | 1263 | | | | | | | |
| Db | 38627 | TTG | 38625 | | | | | | | |

| | |
|------------|---|
| RESULT | 12 |
| ACI07771 | |
| LOCUS | |
| DEFINITION | Mus musculus clone RP23-106C21, LOW-PASS SEQUENCE SAMPLING. |
| ACCESSION | ACI07771.1 |
| VERSION | GI:18308473 |
| KEYWORDS | HTG; HTGS_PHASED. |
| SOURCE | house mouse. |
| ORGANISM | Mus musculus |

| REFERENCE | AUTHORS | TITLE | JOURNAL | REFERENCE |
|----------------------|--|---------------------------------|-------------|-----------|
| 1 (bases 1 to 68591) | Birren, B., Linton, L., Nusbaum, C., and Lander, E. | Mus musculus, clone RP23-106c21 | unpublished | |
| 2 (bases 1 to 68591) | Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, S.,... | | | |

| | |
|---------|--|
| TITLE | Direct Submission |
| JOURNAL | Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA |
| COMMENT | All repeats were identified using RepeatMasker. |

Smit, A.F.A. & Green, P. (1996-1997)
<http://www.1954-1997.nl>

----- Genome Center
 mcp://lcp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR

Web site: <http://www-seq.wi.mit.edu>

----- Contact: sequence_submissions@ncbi.nlm.nih.gov -----

Center project name: L20433

```
-----
Center clone name: 106_C_21
```

* NOTE: This record contains 82 individual

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
714 813: gap of 713 bp in length
814 1570: contig of 100 bp
1571 1670: gap of 100 bp
1671 2427: contig of 757 bp in length
2428 2527: gap of 100 bp
2528 3286: contig of 759 bp in length
3287 3386: gap of 100 bp
3387 4108: contig of 722 bp in length
4109 4208: gap of 100 bp
4209 4945: contig of 737 bp in length
4946 5045: gap of 100 bp
5046 5771: contig of 726 bp in length
5772 5871: gap of 100 bp
5872 6606: contig of 735 bp in length
6607 6706: gap of 100 bp
6707 7430: contig of 724 bp in length
7431 7530: gap of 100 bp
7531 8279: contig of 749 bp in length
8280 8379: gap of 100 bp
8380 9134: contig of 755 bp in length
9135 9234: gap of 100 bp
9235 9954: contig of 720 bp in length
9955 10054: gap of 100 bp
10055 10804: contig of 750 bp in length
10805 10904: gap of 100 bp
10905 11644: contig of 740 bp in length
11645 11744: gap of 100 bp
11745 12458: contig of 714 bp in length
12459 12558: gap of 100 bp
12559 13290: contig of 732 bp in length
13291 13390: gap of 100 bp
13391 14120: contig of 730 bp in length
14121 14220: gap of 100 bp
14221 14959: contig of 739 bp in length
14960 15059: gap of 100 bp
15060 15814: contig of 755 bp in length
15815 15914: gap of 100 bp
15915 16663: contig of 749 bp in length
16664 16763: gap of 100 bp
16764 17506: contig of 743 bp in length
17507 17606: gap of 100 bp
17607 18359: contig of 753 bp in length
18360 18459: gap of 100 bp
18460 19211: contig of 752 bp in length
19212 19311: gap of 100 bp
19312 20036: contig of 725 bp in length
20037 20136: gap of 100 bp
20137 20870: contig of 734 bp in length
20871 20970: gap of 100 bp
20971 21703: contig of 733 bp in length
21704 21803: gap of 100 bp
21804 22550: contig of 747 bp in length
22551 22650: gap of 100 bp
22651 23388: contig of 738 bp in length
23389 23488: gap of 100 bp
23489 24227: contig of 739 bp in length
24228 24327: gap of 100 bp
24328 25079: contig of 752 bp in length
25080 25179: gap of 100 bp
25180 25924: contig of 745 bp in length
25925 26024: gap of 100 bp
26025 26757: contig of 733 bp in length

26758 26857: gap of 100 bp
26858 27608: contig of 751 bp in length
27609 27708: gap of 100 bp
27709 28456: contig of 748 bp in length
28457 28556: gap of 100 bp
28557 29305: contig of 749 bp in length
29306 29405: gap of 100 bp
29406 30140: contig of 735 bp in length
30141 30240: gap of 100 bp
30241 30938: contig of 698 bp in length
30939 31038: gap of 100 bp
31039 31747: contig of 709 bp in length
31748 31847: gap of 100 bp
31848 32577: contig of 730 bp in length
32578 32677: gap of 100 bp
32678 33439: contig of 762 bp in length
33440 33539: gap of 100 bp
33540 34268: contig of 729 bp in length
34269 34368: gap of 100 bp
34369 35100: contig of 732 bp in length
35101 35200: gap of 100 bp
35201 35945: contig of 745 bp in length
35946 36045: gap of 100 bp
36046 36795: contig of 750 bp in length
36796 36895: gap of 100 bp
36896 37624: contig of 729 bp in length
37625 37724: gap of 100 bp
37725 38456: contig of 732 bp in length
38457 38556: gap of 100 bp
38557 39274: contig of 718 bp in length
39275 39374: gap of 100 bp
39375 40119: contig of 745 bp in length
40120 40219: gap of 100 bp
40220 40944: contig of 725 bp in length
40945 41044: gap of 100 bp
41045 41791: contig of 747 bp in length
41792 41891: gap of 100 bp
41892 42644: contig of 753 bp in length
42645 42744: gap of 100 bp
42745 43468: contig of 724 bp in length
43469 43568: gap of 100 bp
43569 44318: contig of 750 bp in length
44319 44418: gap of 100 bp
44419 45168: contig of 750 bp in length
45169 45268: gap of 100 bp
45269 46013: contig of 745 bp in length
46014 46113: gap of 100 bp
46114 46827: contig of 714 bp in length
46828 46927: gap of 100 bp
46928 47632: contig of 705 bp in length
47633 47732: gap of 100 bp
47733 48460: contig of 728 bp in length
48461 48560: gap of 100 bp
48561 49286: contig of 726 bp in length
49287 49386: gap of 100 bp
49387 50127: contig of 741 bp in length
50128 50227: gap of 100 bp
50228 50981: contig of 754 bp in length
50982 51081: gap of 100 bp
51082 51821: contig of 740 bp in length
51822 51921: gap of 100 bp
51922 52658: contig of 737 bp in length
52659 52758: gap of 100 bp
52759 53510: contig of 752 bp in length
53511 53610: gap of 100 bp
53611 54354: contig of 744 bp in length
54355 54454: gap of 100 bp
54455 55195: contig of 741 bp in length
55196 55295: gap of 100 bp
55296 56024: contig of 729 bp in length
56025 56124: gap of 100 bp
56125 56874: contig of 750 bp in length
56875 56974: gap of 100 bp

```

* 56975 57694: contig of 720 bp in length
* 57695 57794: gap of 100 bp
Query Match
Best Local Similarity 4.68; Score 77.6; DB 2; Length 68591;
Matches 134; Conservative 0; Mismatches 74; Indels 2; Gaps 1;

OY 188 cccctcagtcacacgactcctcctcctcgcgcacaaagtcacaaagaagcctaac 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51568 ccccccggcccccggcccccggcccccggcccccggcccccggcccccggccccc 51627
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 248 aggaagcgaaggaaggaaggaaggaaggaagcagcagcagcagcagcagcagcagc 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51628 aggcagcgaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 51687
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 308 tgcgggaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 367
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51688 tgcagagatagctctg--gcagatctcccatctgagacagcctctgacacacgctctgag 51745
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 368 gcaatcattctgctgaaggtcaccagc 397
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51746 atacatcatatgtgtgagtgagtgccgcacagc 51775
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 13

```

AC094188 89536 bp DNA linear HTG 20-DEC-2001
LOCUS Rattus norvegicus clone CH230-2K16, *** SEQUENCING IN PROGRESS ***
DEFINITION 49 unorderd pieces.
ACCESSION AC094188
VERSION AC094188.2 GI:17940901
KEYWORDS HTG, HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.

```

REFERENCE

AUTHORS

```

1 (bases 1 to 89536)
Muzny,D.M., Adams,C., Adio-Ogboia,B., Ali-Osman,F.R., Allen,C.,
Albrooks,S.L., Amaralungue,H.C., Are,J.R., Banks,T., Barbieri,J.,
Benton,J., Bimberg,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Bivlev,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douhwalte,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flaggs,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C.,
Hollins,B., Homsj,P., Howard,S., Huber,J., Huliy,S., Hume,J.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvath,J.,
Kovar,L., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lounsged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,D., Mitchell,T., Mohabbat,K.,
Moran,M., Morris,S., Moser,M., Neal,D., Newton,J., Newson,N.,
Nguyen,A., Nguyen,N., Oragune,N., Nickerson,E., Nwokenko,S.,
Ogun,M., Okunoye,G., Oragune,N., Overdone,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Prins,E., Pu,L.L.,
Ruijs,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tameris,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,

```

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```

Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gdbbs,R.
Direct Submission
Unpublished
2 (bases 1 to 89536)
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15636537.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GABR
Center clone name: CH230-2K16
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 73144 bases at least Q40
Consensus quality: 80017 bases at least Q30
Consensus quality: 85368 bases at least Q20
Estimated insert size: 69357; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation
-----
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 49 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

```

```

1 4713: contig of 4713 bp in length
4714 4813: gap of unknown length
4814 7442: contig of 2629 bp in length
7443 7543: gap of unknown length
7543 8970: contig of 1428 bp in length
8971 9071: gap of unknown length
9071 11618: contig of 2548 bp in length
11618 11719: gap of unknown length
11719 14064: contig of 2346 bp in length
14064 14164: gap of unknown length
14164 14659: contig of 2434 bp in length
14659 16598: gap of unknown length
16598 16988: gap of unknown length
16988 18558: contig of 1860 bp in length
18558 18658: gap of unknown length
18658 21636: contig of 2978 bp in length
21636 21737: gap of unknown length
21737 23091: contig of 1355 bp in length
23091 23191: gap of unknown length
23191 24363: gap of 1172 bp in length
24363 24463: gap of unknown length
24463 26616: gap of 2153 bp in length
26616 26716: gap of unknown length
26716 29328: contig of 2612 bp in length
29328 29428: gap of unknown length
29428 30861: contig of 1433 bp in length
30861 30962: gap of unknown length
30962 32667: contig of 1705 bp in length
32667 32767: gap of unknown length
32767 34751: gap of unknown length
34751 34852: contig of 1985 bp in length
34852 36722: contig of 1871 bp in length
36722 36822: gap of unknown length

```


Matches 76; Conservative 284; Mismatches 275; Indels 8; Gaps 2;

| | | | |
|----|-------|--|-------|
| OY | 60 | gagggagagctccaagcagctgcccagcagcttgcagatccctgcgaacttgcctccctctgaacctg | 119 |
| Db | 17148 | KKSMKCYCTYMSMKGYYACYSRGSMSKTCMRGISTYSTGGCCCTTTTTCGCCNN | 17207 |
| OY | 120 | gaagatgycggcccgctgtagcttaacctctcgtgtgttgcgtgcgaacctgagt | 179 |
| Db | 17208 | TGGGGAGCTTTTNCNKRTYSYRRRNGCAMKYNINYNSMRSSCRAGMSTCKTYSMT | 17267 |
| OY | 180 | ccaagccaccttcagtcacctcagctgcagttctcaatcctcggcccaaaagctaaagaaa | 239 |
| Db | 17268 | MSMASYCWCMASYTCSMRMSARGSWSWSSYMMKMSWPMSCYMKCMCMKCMYCMRSM | 17327 |
| OY | 240 | gctgcacacagagcgtcgaagagaccacaagccacccagcaccctcgacagcagctgcgctgt | 299 |
| Db | 17328 | RSRGMSTYYTASWKKSSSRKCTCTTCYWCSSKSTCKSTYMMNRSKRMKMSRCWGSNW | 17387 |
| OY | 300 | cagtcgcatctgcgggaaaagccagccggagatccctctgctgcggagcctgtgtgaacacc | 359 |
| Db | 17388 | GSASRSSCYKYCKSMRCSWMSKCYRCAGCMNMKGMYMCMGMRKRMWCKSLMYRR | 17447 |
| OY | 360 | gtccctgaagcaatcatctctgctgtaagtgatcaatcaagctaacaatcctcagctgaagtg | 419 |
| Db | 17448 | RMWRGAMMCMKCYSRMSRCMMMKSYCACRCAMMSGTYMCASTMCSSWTCTW | 17507 |
| OY | 420 | aagccctcgccaatgacacagagcgtgctgaagtcaga--tcccctgacatagctgctg | 477 |
| Db | 17508 | CTSYMSYRCTCMCGWGSYTKSKMSMSSMSYTKKGRKRSYMCCTSRAMSCMRCTCM | 17567 |
| OY | 478 | gattcaacaagccctcgtgtcaagaccatcgtgtgagttccaatgaaga----ctgag | 531 |
| Db | 17568 | RGASSMRAGSMRRRAKRSRSMWGRSKMMWTGCMRSKYYTCTGRMMMTYCCCMRRRS | 17627 |
| OY | 532 | ccaagaccacaatccgcgaatgacacacagctgaagcggcccccacccgctgctcccatg | 591 |
| Db | 17628 | YMYTSMNAGMKRKSNSGMRMGASARRCKSARSSMCSRRMKGMGMSWSSMMWGS | 17687 |
| OY | 592 | actgtgcacaagcacaatgagagcctgcgcatccaactcgtcatlaagacttctctctcgtg | 651 |
| Db | 17688 | RSRSASMKCSKSGRRRRSRKSSRYKRGKRRKMTKSKGSKGKCMKWRGSMTSSCY | 17747 |
| OY | 652 | tgaagccttagctaaagcagtgtaagaactcctcctagctcattc 694 | |
| Db | 17748 | YYSASSCMHMSKSCMCCMMKRCACCTYSMSCTSMYRCC 17790 | |

RESULT 15
BTU79413
LOCUS BTU79413
qqq bp
EBNA 1 repeat
MAR 00-DEC-1006

```

CDS
      /db_xref="taxon:9913"
      /tissue_type="parotid gland"
      36..767
      /note="Common salivary protein, form a"
      /codon_start=1
      /product="BSP30"
      /protein_id="AAB38282.1"
      /db_xref="GI:1710367"
      /translation="MVQIMKLVLLGLAGTSSELDIRGMDVLRRLISGLERLIGTFE
      DSTLEILFQNLKTELESKLDVVEEQQTENSLGLISRFQVNNRLTGVRINVOV
      VVIFFEATSENSADVSPITADAVNSPLGLIEIVKLDLVNQLQTSVSIETDAEGDSK
      PVVEECNPNPESISLVFLHRPGLINDVDEGVNMLYRQLVASVQHEQLPRIRLELSS
      LDTECIKKLIEPQVTLQQQSESVNR"
BASE COUNT      267 a      247 c      255 g      230 t
ORIGIN

```

| Query Match | 3.2% | Score 54.6 | DB 4 | Length 999 |
|-----------------------|----------------|--|----------|------------|
| Best Local Similarity | 50.5% | Pred. No. 0.02% | | |
| Matches 160 | Conservative 0 | Mismatches 154 | Indels 3 | Gaps 1 |
| QY | 450 | gtcaagaatcccccttgaacatgtgtgtcgtg9attccaacacgcccctgtgtaagaacatcgtg | 509 | |
| Db | 384 | ggtgtgatgtcccatcattacacgcgtgacgtcaccgctgacccgttccctgttgggtgagattgttc | 443 | |
| QY | 510 | gaattccacatgaagacttgaagtgaggcccaagccacatccgcgatggaacaccagtgc---aagt | 566 | |
| Db | 444 | aaactggagactctcaattgtaggacctccaacttactgtctcagcatfgaamaacagatccttgagact | 503 | |
| QY | 567 | ggccccacccgcctgtgtctcctcagtgacgtgtgtgcacacagcatcgggaagcttgcgcatccaa | 626 | |
| Db | 504 | ggtgactctcaggggtgtgtgtggagagaaatgcccccaacaaacccagaaagacattctcactcagc | 563 | |
| QY | 627 | ctgtgtgcaaaagcttccttcctgtgtgaagcgtttagctlaagcagtgatgaacctctcta | 686 | |
| Db | 564 | gtgtgtgcacagggccgccttgaaactgcttcaacgattgttggagattttggagctcaaccttga | 623 | |
| QY | 687 | gtgtccatcccttgcgccaatctagtgaaaaaacagctgtgttcccgtatcgtagagcttccttc | 746 | |
| Db | 624 | agacagtttggtgtctctctgtgtagtcgacgacgagctgtgcccacgaatccgtgaactcttt | 683 | |
| QY | 747 | aatggcatgtatgacaga | 763 | |
| Db | 684 | gaaaaccttgatpacaga | 700 | |

Search completed: September 16, 2002, 02:53:46
Job time: 7439 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:28:35 ; Search time 3813.96 Seconds

(without alignments)
1728.350 Million cell updates/sec

Title: US-09-700-770-4

Perfect score: 315
Sequence: 1 taaactgctcgaattt.....tggatattacttattatgc 315

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| ----- | | | | | |

| | | | | | | | |
|---|----|-----|-------|--------|----|------------|--------------------|
| C | 1 | 315 | 100.0 | 118068 | 9 | AL356798 | AL356798 Human DNA |
| C | 2 | 21 | 6.7 | 3705 | 14 | NPHEXA | M16821 Autographa |
| C | 3 | 21 | 6.7 | 133694 | 6 | A48542 | A48542 Sequence 1 |
| C | 4 | 21 | 6.7 | 133894 | 14 | I22858 | L22858 Autographa |
| C | 5 | 21 | 6.7 | 142289 | 9 | AC078955 | AC078955 Homo sapi |
| C | 6 | 21 | 6.7 | 163283 | 9 | AL355606 | AL355606 Human DNA |
| C | 7 | 20 | 6.3 | 57287 | 9 | AC008283 | AC008283 Homo sapi |
| C | 8 | 20 | 6.3 | 81580 | 8 | AP000736 | AP000736 Arabidops |
| C | 9 | 20 | 6.3 | 87966 | 2 | AC105559 | AC105559 Rattus no |
| C | 10 | 20 | 6.3 | 98017 | 8 | AC027033 | AC027033 Arabidops |
| C | 11 | 20 | 6.3 | 109881 | 2 | AC096828 | AC096828 Rattus no |
| C | 12 | 20 | 6.3 | 111159 | 9 | AC008271 | AC008271 Homo sapi |
| C | 13 | 20 | 6.3 | 114927 | 2 | AC105629 | AC105629 Rattus no |
| C | 14 | 20 | 6.3 | 11548 | 2 | AC094962 | AC094962 Rattus no |
| C | 15 | 20 | 6.3 | 117757 | 9 | HSU100419 | AL121912 Human DNA |
| C | 16 | 20 | 6.3 | 139619 | 2 | AL627164 | AL627164 Danio rer |
| C | 17 | 20 | 6.3 | 146950 | 2 | AC095100 | AC095100 Rattus no |
| C | 18 | 20 | 6.3 | 152695 | 2 | AC095576 | AC095576 Mus muscu |
| C | 19 | 20 | 6.3 | 161394 | 2 | AC010759 | AC010759 Homo sapi |
| C | 20 | 20 | 6.3 | 211681 | 2 | AC083915 | AC083915 Mus muscu |
| C | 21 | 19 | 6.0 | 512 | 11 | AF220278 | AF220278 Oncorhyn |
| C | 22 | 19 | 6.0 | 912 | 8 | AB011794 | AB011794 Citrus si |
| C | 23 | 19 | 6.0 | 1042 | 10 | S78689 | S78689 LBL-1=large |
| C | 24 | 19 | 6.0 | 5313 | 8 | SCYNR054C | Z71670 S.cerevisia |
| C | 25 | 19 | 6.0 | 7310 | 8 | SCYNR055C | U39740 Caenorhabdi |
| C | 26 | 19 | 6.0 | 36148 | 3 | U39740 | AL354726 Human DNA |
| C | 27 | 19 | 6.0 | 82316 | 9 | AL354726 | AL157759 Human DNA |
| C | 28 | 19 | 6.0 | 83552 | 9 | AL157759 | AC009540 Arabidops |
| C | 29 | 19 | 6.0 | 101410 | 8 | ATAC009540 | AC097581 Rattus no |
| C | 30 | 19 | 6.0 | 125245 | 2 | AC097581 | AC108885 Rattus no |
| C | 31 | 19 | 6.0 | 138328 | 2 | AC108885 | AC025173 Homo sapi |
| C | 32 | 19 | 6.0 | 140935 | 9 | AC025173 | AC097721 Homo sapi |
| C | 33 | 19 | 6.0 | 144493 | 9 | AC097721 | AC080176 Homo sapi |
| C | 34 | 19 | 6.0 | 145496 | 2 | AC080176 | AC006063 Homo sapi |
| C | 35 | 19 | 6.0 | 150001 | 9 | AC006063 | AC048373 Homo sapi |
| C | 36 | 19 | 6.0 | 152631 | 2 | AC048373 | AC108150 Homo sapi |
| C | 37 | 19 | 6.0 | 154307 | 2 | AC108150 | AC093250 Homo sapi |
| C | 38 | 19 | 6.0 | 157332 | 2 | AC093250 | AC098610 Rattus no |
| C | 39 | 19 | 6.0 | 161059 | 2 | AC098610 | AC096721 Homo sapi |
| C | 40 | 19 | 6.0 | 161824 | 9 | AC096721 | AC009662 Homo sapi |
| C | 41 | 19 | 6.0 | 175448 | 2 | AC009662 | AC097514 Homo sapi |
| C | 42 | 19 | 6.0 | 177027 | 9 | AC097514 | AP002378 Homo sapi |
| C | 43 | 19 | 6.0 | 177429 | 9 | AP002378 | AC023884 Homo sapi |
| C | 44 | 19 | 6.0 | 178310 | 2 | AC023884 | AC010763 Homo sapi |
| C | 45 | 19 | 6.0 | 180185 | 2 | AC010763 | |

ALIGNMENTS

RESULT 1
LOCUS AL356798/c
DEFINITION Human DNA sequence from clone RP11-467B11 on chromosome 9, complete sequence.
ACCESSION AL356798
VERSION AL356798.18 GI:15795413
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 118068)
AUTHORS Babbage,A.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Sep 27, 2001 this sequence version replaced gi:14455883.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, SW, SWISSPROT, Tr, TREMBL, Wp, WORMPEP; information on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
Rp11-467B11 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rp11-467B11. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone Rp11-467B11 is at 118068 in this sequence. The true left end of clone Rp11-208P1 is at 29578 in this sequence. The true right end of clone Rp11-29B11 is at 2000 in this sequence.

FEATURES

source
1..118068
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="Rp11-467B11"
/clone_id="RPC1-11.2"
BASE COUNT 33053 a 23625 c 22510 g 38880 t
ORIGIN

Query Match 100.0%; Score 315; DB 9; Length 118068;

Best Local Similarity 100.0%; Pred. No. 1.6e-157;

Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taacacgtacacgatttaagaataactttgagaataagaacaatagaatcagtt 60
Db 96845 TAAACAGCATCAGATTTTAAGAAATTAAGTTGAGAAATAGAAACAATGAATCAGTT 96786
QY 61 tctccaccacttaagatatactcttagagatactacagccctcccttagggagatacaaa 120
Db 96785 TCTCCACCACCTTAAGTATCTCTTAAAGATCTACAGCCCTCTTAAAGGAGATCAAA 96726
QY 121 gtccagttgttgccttggtagtccaccattatcgaagttagatgatacaaat 180
Db 96725 GTCCAGTTGTGCTTGGTGTGAGTCCACCTTATATTCAGTAGTATGACTCAAAAT 96666
QY 181 ttgaagaatagatgtcacacaataaactgaggtttatgagaacatcagtagagaataa 240
Db 96665 TTGAAATATAGATTGTACACAAATAACTGAGATTATGGAACATCAGTAGAAGAAATA 96606
QY 241 caacattccatccctttagagagatacttctgcaactcaggagaattgtcaltgt 300
Db 96605 CAACATTCATCCCTTTAAGAGATATTACTTGCAACTCAGGATAATTTGTCTAGTGT 96546
QY 301 attatctactatgc 315
Db 96545 ATTATCTACTTATGC 96531

RESULT 2
NPHKEA/c 3705 bp DNA linear VRL 07-FEB-1994
LOCUS NPHKEA
DEFINITION Autographa californica nuclear polyhedrosis virus HindIII K region

ACCESSION
VERSION M16821.1 GI:332445
KEYWORDS 35 kDa protein; 94 kDa protein.
SOURCE Autographa californica nuclear polyhedrosis virus (strain L-1) DNA, passed in Spodoptera frugiperda (IPB-SF-21 cells).
ORGANISM Autographa californica nucleopolyhedrovirus
Virus; dsDNA viruses, no RNA stage; Baculoviridae; Nucleopolyhedrovirus.

REFERENCE
AUTHORS Friesen, P.D. and Miller, L.K.
TITLE Divergent transcription of early 35- and 94-kilodalton protein genes encoded by the HindIII K genome fragment of the baculovirus Autographa californica nuclear polyhedrovirus
J. Virol. 61, 2264-2272 (1987)
87226411

COMMENT
Displaced messenger-active RNAs are transcribed from both 35- and 94-kD protein genes early (2 hours) after infection. Late in infection, abundant RNAs are transcribed from promoters located at least 2.5 kb upstream from the gene encoding the 35 kD protein. These transcripts completely overlap both the 35 and 94 kD polypeptide genes, but apparently lack protein-coding potential, suggesting that the transcripts may play a role in suppressing early viral gene expression.

FEATURES

source

1..3705
/organism="Autographa californica nucleopolyhedrovirus"
/db_xref="taxon:46015"
complement(39..2684)
/note="early 94 kb mRNA (alt.)"

mRNA
complement(39..2503)
/note="early 94 kb mRNA (alt.)"

mRNA
complement(39..2569)
/note="early 94 kb mRNA (alt.)"

CDS
complement(79..2490)
/note="early 94 kb protein"

/codon_start=1
/protein_id="AAA46702.1"

/db_xref="GI:332446"

/translation="MINTVYATDSDSTNSDYFNKNALQTLEFQNEIENISSCDK
VLIHMSVACQKELGDKVSKYLRNGEGDTPSEFIKWIKHKNILNIDSKILVY
TDGQISKNEAVCNKLNLEKPSFERIVFYVINNTQIDLSVASAVNNSSDKIYRN
DEWVWVNI/KEFNVDITLTENFISKDELLSFREFEINSPMDAVINENYKIKTL
RORLESEIKOTNNSMNPDOIKKNEVEVPEKSEFEYKTYINADVLEPDTIDSTST
ALINLHNKRSYAFDVMKNLHYOKKLSVNAVEDVYANDDEAYDYSNENRFPDC
ILANDSGYPATILLHYNLPETIGSLTRFSRDEPLVMSQNEIKNSIECYNLESL
KQIIOHGRISPSRRPPTGAIVNEQFDENDVLACTFYDAKKAFFNAGMYLILY
KHINDATIDNVADYFRKRYIVRINTECOMIGSNLAMEPLIKVLPALMYSEIS
TLERKHNQHFGEKELRQFAHFAEDMIQIOWMCDYDVNVAVKRAYCLKRIIMFER
MSVDAWEMIANRAFECKDEKFIINKLNADALQDLKLVNHHGVNDEHYVAVLKGEF
ERYLYFVHIIEDEPKYISVDIMRPAFVLEKKTPEYDILKQLOSYPHNGQITTEK
CSRLEFNRIKLSHRLYTECVKSLNKYPTLECYONYYNQKHVFNKRIAPPEITQNL
AAVHNEYANKIVNLPVEEIVRANRYNRIITRIONEVSPLOAEEDIKLIKLEQRY
NICRK"

mRNA
2667..3607
/note="early 35 kb mRNA (3' end +/- 3 bp)"

CDS
2701..3600
/note="early 35 kb protein"

/codon_start=1
/protein_id="AAA46703.1"

/db_xref="GI:332447"

/translation="KCVIFPVEIDVSTIIRDCQVDKQRELVYINKINNTQITKPV
LMPNIGSPINRSVTRKNNNNLRKISVDEDFQLEBDYSDOMGFDHSIKYKDEHS
VSCQNGSVLSKSPAKTIKSHDYDDEKSIKSAVEYCLPKLVDERNDYVAVVCLKGEF
NGSNQVLSFEYFNIGNKYIVPEAHENIDNGLEYDVAVVDSVQDGBQDFEYFQSLI
LPSSFKNSEKVLVYNESKNSKMYTALAEFTTBESSGKSEKIMKIFCGEFLYDKSK
VLVYKLAHNTSALNKVILNTIK"

BASE COUNT 1080 a 686 c 731 g 1208 t

ORIGIN 122 bp upstream of HindIII site; 85-88 mu on the viral map.

Query Match 6.7%; Score 21; DB 14; Length 3705;
Best Local Similarity 100.0%; Pred. No. 2.5;

Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 aagaataactttgagaat 41
|||||
Db 464 AAGAAATACTTTGAGAAAT 444

RESULT 3
AA8542/c 133894 bp DNA linear PAT 07-MAR-1997
LOCUS Sequence 1 from Patent WO9601320.
DEFINITION A48542
ACCESSION A48542
VERSION A48542.1 GI:2302312
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 133894)
AUTHORS Bishop,D., Possee,R. and Ayres,M.
TITLE AUTOGRAPHA CALIFORNICA COMPLETE GENOME SEQUENCE
JOURNAL PATENT: WO 9601320-A 1 18-JAN-1996;
NATURAL ENVIRONMENT RES (GB)
COMMENT Other publication AU 2897295 960125.
FEATURES
source location/Qualifiers
1..133894
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 39195 a 27151 c 27347 g 40201 t
ORIGIN

Query Match 6.7%; Score 21; DB 6; Length 133894;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 aagaataactttgagaat 41
|||||
Db 114255 AAGAAATACTTTGAGAAAT 114235

RESULT 4
L22858/c 133894 bp DNA circular VRL 29-MAR-2001
LOCUS Autographa californica nucleopolyhedrovirus clone C6, complete
DEFINITION genome.
ACCESSION L22858
VERSION L22858
KEYWORDS L22858.1 GI:510708
SOURCE Autographa californica nucleopolyhedrovirus.
ORGANISM Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
REFERENCE 1 (bases 1 to 133894)
AUTHORS Ayres,M.D., Howard,S.C., Kuzio,J., Lopez-Ferber,M. and Possee,R.D.
TITLE The complete DNA sequence of Autographa californica nuclear
polyhedrosis virus
JOURNAL Virology 202 (2), 586-605 (1994)
MEDLINE 94303173
PUBMED 8030224
REFERENCE 2 (bases 1 to 133894)
AUTHORS Kuzio,J.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-1999) NCBI, Bethesda, MD 20894, USA
FEATURES
source location/Qualifiers
1..133894
/organism="Autographa californica nucleopolyhedrovirus"
/db_xref="taxon:46015"
/clone="C6"
503..1009
/gene="Ac-ptp"
/note="Acorf-1"
503..1009
CDS
/gene="Ac-ptp"

/note="PTP; 19288 Da primary translation product"
/codon_start=1
/product="protein tyrosine phosphatase"
/protein_id="AA66631.1"
/db_xref="GI:559070"
/translation="MEPARHNYVLOCGGVKIDSNLCIKRTPLRPFLFAVYVSEEDVMT
AEQIKONPSIGAILDINTSKSYVDGVHFKGLYKKIQVGGTLPPESIVQEFIDT
VKEFEKCPGMVGVCHTNGINRTGYMCRYLMHTLGIAPEADIDREFKAGHKIERQ
NYVDLLT"
complement(1041..2027)
/gene="Ac-bro"
/note="Acorf-2"
complement(1041..2027)
/gene="Ac-bro"
/note="BRO; 37769 Da primary translation product"
/codon_start=1
/product="baculovirus repeated ORF"
/protein_id="AAA6632.1"
/db_xref="GI:559071"
/translation="MARVKIGFEKFGEDPNTLRVYLERDQGVREYAKVANSIKYTVG
DKATRVHDKYKSLFEOTIONGPSTSVYKRGDPLQHPVLTITSGVQLINKS
KLPAIEIQLLEEVITPOVLCTGTIDPAIKOREESQVLTATTEHTNALQAV
VAQTEELVKRQEFLETVAIKDKQIEAKDLQVTRVMTDLNMYGFEQEMOKDEIM
OKDKAVDVLAKVVDLSDRAVOYPADRKRHVLCTVDGTFTEATIGQKTYVENQKH
KRNIIVANIVENIRPNPTVDNMNATDRLQAKRSRSTIVLWKKRRNLKIG"
complement(2084..2245)
/gene="Ac-ctx"
/note="Acorf-3"
complement(2084..2245)
/gene="Ac-ctx"
/note="CTX; 5590 Da primary translation product"
/codon_start=1
/product="conotoxin-like peptide"
/protein_id="AAA6633.1"
/db_xref="GI:559072"
/translation="MQIKTVLLAFAMFALNAQHVLAACETGAVCVHNDCCGAGCS
PIFNCLPQ"
complement(2295..2750)
/gene="Acorf-4"
complement(2295..2750)
/gene="Acorf-4"
/note="17577 Da primary translation product"
/codon_start=1
/product="Acorf-4 peptide"
/protein_id="AAA6634.1"
/db_xref="GI:559073"
/translation="MKLYTKASLSKVALRTRKENTIPHFILTRRLDIDGMIX
SEDVGFENRTNRNDLISACMQIIVRTYMPNATIDMRKQNPCTYFRICQCHLEADVPS
PDDHSVYRILCVACCTPIVDHPLDVFHTEGVELLEVORVAGGEL"
2779..3108
/gene="Acorf-5"
2779..3108
/gene="Acorf-5"
/note="12435 Da primary translation product"
/codon_start=1
/product="Acorf-5 peptide"
/protein_id="AAA6635.1"
/db_xref="GI:559074"
/translation="WRTSRINNAPEVVASQHDYRDQIKRELNSLRVNHVICTRSGT
SEFDCNPLRSDMTVPVTTIPKRTADYKITEYVGDVATIKRNPPLVESGLVREAA
KYGECIV"
3089..3721
/gene="Ac-lef2"
/note="Acorf-6"
3089..3721
/gene="Ac-lef2"
/note="LEF2; 23926 Da primary translation product"
/codon_start=1
/product="late expression factor 2"
/protein_id="AAA6636.1"
/db_xref="GI:559075"
/translation="MANASYNWSPILIRASCLDKRATYVLDPPDFIDRLTLTPYTFY
NGVLVTKISGLRMLTAPPTINIKSNFRRKRNKCMKECEVGEKKNVDMLNK

[illegible]

Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenye, S., Ogun, M., Okunolu, G., Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojudo, K., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shoshitaishvili, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Zuchner, R., and Gibbs, R.

Unpublished
2 (bases 1 to 142289)
Worley, K.C.
Direct Submission
Submitted (13-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 142289)
Worley, K.C.
Direct Submission
Submitted (30-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 30, 2001 this sequence version replaced gi:13470106.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc_help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 142289
Phrap values in estimate: 141529
Average error rate (BCM-Phrap estimate): 1.31902e-05

Fraction of Phrap values less than 40 : 0.00705862
Number of consensus changing edits: 41
Number of N's in consensus : 0

| Position | Original+Context | Edited+Context |
|----------|--------------------------|--------------------------|
| 51 | aacattatg(n)attttttg | aacattatg(v)attttttg |
| 62 | attttttg(n)ttttttgt | attttttg(t)ttttttgt |
| 67 | ttttttttg(n)ttgttttt | ttttttttg(t)ttgttttt |
| 3797 | taatttaagt(n)aaccaccaaa | taatttaagt(a)aaccaccaaa |
| 41447 | ctccctccctc(n)acatgacac | ctccctccctc(a)acatgacac |
| 41447 | gtacatgag(n)anncggttgc | gtacatgag(a)anncggttgc |
| 41448 | taagctagag(n)gtttanaaaa | taagctagag(t)gtttanaaaa |
| 41449 | agcttagagan(n)mcggttgt | agcttagagat(a)mcggttgt |
| 41450 | gctagagan(n)cggttgtgc | gctagagat(a)cggttgtgc |
| 41631 | ltaagaaga(n)accacagggc | ltaagaaga(g)accacagggc |
| 43314 | ctctgtactg(n)acatgacag | ctctgtactg(t)acatgacag |
| 68569 | gtgataagct(n)gtttanaaaa | gtgataagct(a)gtttanaaaa |
| 68571 | gataagact(n)ltaanaaat | gataagact(a)ltaanaaat |
| 68576 | actngtttaa(n)aaatgtgtgc | actagattaa(g)aaatgtgtgc |
| 68587 | aaatgtgtgc(n)catntttcc | aaatgtgtgc(a)catntttcc |
| 68591 | ttgtgmcac(n)ttttccatgg | ttgtgacacat(a)ttttccatgg |
| 68593 | ltygcncatnt(n)ttccatggaa | ltygcacacat(a)ttccatggaa |
| 68595 | gncatntnt(n)ccatggaatn | gacacatata(a)ccatggaatn |
| 68605 | ncatggaat(n)ttatgcaccc | accatggaat(a)ttatgcaccc |
| 68606 | ccatggaat(n)tatgcagcca | ccatggaat(a)tatgcagcca |
| 68742 | ttggaattga(n)caatgagatc | ttggaattga(a)caatgagatc |
| 74888 | gagactccgt(n)tcnaaaaaa | gagactccgt(c)tcnaaaaaa |
| 74891 | actccgtntc(n)aaaaaaaaa | actccgtntc(a)aaaaaaaaa |
| 75765 | lgaataaaa(n)agatatacaa | lgaataaaa(g)agatatacaa |
| 77345 | ltycagcacac(n)nnntnnnac | ltycagcacac(c)nnntnnnac |
| 77346 | gcagcacacn(n)nnntnnnac | gcagcacacac(a)nnntnnnac |
| 77347 | cagcacacnn(n)ntnnnacat | cagcacacac(a)ntnnnacat |
| 77348 | agcacacnn(n)ntnnnacat | agcacacac(a)ntnnnacat |
| 77349 | gcacacnnn(n)ttnnnacatg | gcacacacaa(c)ttnnnacatg |
| 77351 | gcacacnnn(n)nnacatgata | gcacacacaa(a)nnacatgata |
| 77352 | ccacnnntn(n)nnacatgata | ccacacacat(g)nnacatgata |
| 77353 | acnnnnntn(n)nnacatgata | acacacacat(g)nnacatgata |
| 77389 | gcacgttgtg(n)nnatgttccc | gcacgttgtg(c)nnatgttccc |
| 77390 | caagtttgtg(n)natgttccc | caagtttgtg(c)natgttccc |
| 77391 | acgttgtg(n)atgttccccta | acgttgtg(c)atgttccccta |
| 77396 | gtgmnatgt(n)cccttaaaat | gtgacatgt(a)cccttaaaat |
| 92451 | aacctaaatt(n)naattgtgac | aacctaaatt(t)naattgtgac |
| 92452 | acctaaatt(n)naattgtgaca | acctaaatt(g)naattgtgaca |
| 109950 | agtgaaataa(c)ccagacagag | agtgaaataa(a)ccagacagag |
| 119939 | ttctgagttt(n)ctcagaataa | ttctgagttt(t)ctcagaataa |
| 120003 | acctggaaca(n)tcagcacctg | acctggaaca(g)tcagcacctg |

----- Distribution of Quality < 40 Bases -----

| # | Phrap Value Range | |
|------|-------------------|---|
| 1000 | 5 | |
| 900 | 10 | |
| 800 | 15 | |
| 700 | 20 | |
| 600 | 25 | |
| 500 | 30 | * |
| 400 | 35 | * |
| 300 | 40 | * |
| 200 | | * |
| 100 | | * |
| 0 | | * |

----- FEATURES -----
Version: 1.01 qxf0.
Location/Qualifiers 1. 142289
source /organism="Homo sapiens"

/db_xref="taxon:9606"
/chromosome="12q"
/clone="RP11-543H12"
1..2000
/note="Overlaps bases 42818..44817 of clone AC004550"
/function="Overlaps with adjacent clone AC004550"
repeat_region
53..80
/rpt_family="TTTTGn"
complement(81..362)
/rpt_family="AluY"
repeat_region
complement(597..967)
/rpt_family="L2"
repeat_region
1019..1811
/rpt_family="L1PA16"
3063..3347
repeat_region
/rpt_family="AluSg"

Query Match 6.7%; Score 21; DB 9; Length 142289;
Best local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 gaatgacaatgaatca 57
|||||
Db 119773 GAAATGAGACAAATGAATCA 119793

RESULT 6
AL355606/c 163283 bp DNA linear PRI 09-MAR-2001
LOCUS Human DNA sequence from clone RP11-9017 on chromosome X, complete
DEFINITION
AL355606
VERSION
AL355606.14 GI:13273693
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNML
Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Mar 11, 2001 this sequence version replaced gi:12743783.
Requests: clonerequest@sanger.ac.uk

COMMENT
Direct Submission
During difference assembly data is compared from overlapping clones.
Where difference are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/projects/c-elegans/wormPep
This sequence was generated from part of bacterial clone configs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChX
RP11-9017 is from the library RP11-11.1 constructed by the group of
Pieret de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-9017.

FEATURES

Location/Qualifiers

source
1..163283
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-9017"
/clone_11b="RP11-11.1"
1004..1071
/note="2 copies 34 mer 89% conserved"
2756..3638
/note="CpG island"
/evidence="not_experimental"
3737..3816
/note="L2 repeat: matches 2100..2179 of consensus"
3861..3983
/note="L2 repeat: matches 2049..2180 of consensus"
3975..4059
/note="MIR repeat: matches 15..102 of consensus"
4060..4360
/note="AluSx repeat: matches 1..303 of consensus"
4361..4534
/note="MIR repeat: matches 102..262 of consensus"
4559..4694
/note="L2 repeat: matches 2208..2355 of consensus"
5091..5388
/note="AluSx repeat: matches 1..299 of consensus"
6254..6612
/note="MER67C repeat: matches 47..424 of consensus"
6719..7019
/note="L2 repeat: matches 2217..2500 of consensus"
7049..7515
/note="MER31B repeat: matches 1..485 of consensus"
7755..8058
/note="AluY repeat: matches 1..309 of consensus"
8421..9284
/note="L2 repeat: matches 902..1813 of consensus"
9346..9630
/note="AluSg repeat: matches 1..293 of consensus"
9794..9919
/note="L2 repeat: matches 2626..2750 of consensus"
10518..10971
/note="L2 repeat: matches 1832..2298 of consensus"
11443..11738
/note="L2 repeat: matches 1832..2298 of consensus"
12805..13029
/note="AluY repeat: matches 1..297 of consensus"
13148..13910
/note="AluY repeat: matches 86..311 of consensus"
13907..14705
/note="L1PA16 repeat: matches 5383..6157 of consensus"
14704..14957
/note="L1PA7 repeat: matches 5347..6143 of consensus"
14958..15248
/note="L1P repeat: matches 4046..4299 of consensus"
15240..15575
/note="L1P repeat: matches 4603..4893 of consensus"
15572..15938
/note="L1P repeat: matches 4285..4621 of consensus"
19940..20174
/note="L1P3 repeat: matches 1..4059 of consensus"
20181..20208
/note="L1P repeat: matches 5155..5398 of consensus"
21901..22068
/note="tRNA-Gly-GCC repeat: matches 1..28 of consensus"
22727..22780
/note="MT1J repeat: matches 1..165 of consensus"
23407..23594
/note="L2 repeat: matches 2699..2750 of consensus"
23601..23767
/note="MIR repeat: matches 48..256 of consensus"
24307..24871
/note="MER5A repeat: matches 4..153 of consensus"
25135..25364
/note="MT1B repeat: matches 1..553 of consensus"
/note="MT1J repeat: matches 109..319 of consensus"

| | | |
|---------------|--------------|---|
| repeat_region | 25433..25505 | /note="MIR11 repeat: matches 319..402 of consensus" |
| repeat_region | 25872..25911 | /note="20 copies 2 mer tt 80% conserved" |
| repeat_region | 25913..26066 | /note="AluSg/x repeat: matches 156..312 of consensus" |
| repeat_region | 26170..26464 | /note="AluSc repeat: matches 17..309 of consensus" |
| repeat_region | 26822..27035 | /note="MIR repeat: matches 4..228 of consensus" |
| repeat_region | 28226..28682 | /note="L1R40a repeat: matches 68..519 of consensus" |
| repeat_region | 28683..29029 | /note="L1Mc repeat: matches 274..615 of consensus" |
| repeat_region | 29055..30453 | /note="L1MA4 repeat: matches 4900..6294 of consensus" |
| repeat_region | 30454..31023 | /note="L1Mc repeat: matches 608..1185 of consensus" |
| repeat_region | 31038..31154 | /note="L2 repeat: matches 2623..2739 of consensus" |
| repeat_region | 31364..31448 | /note="H1nger3(Golem) repeat: matches 1..92 of consensus" |
| repeat_region | 31455..34275 | /note="L1PAs repeat: matches 3298..6145 of consensus" |
| repeat_region | 34276..34571 | /note="AluY repeat: matches 1..296 of consensus" |
| repeat_region | 34572..35031 | /note="L1PAs repeat: matches 2840..3298 of consensus" |
| repeat_region | 35039..35092 | /note="MER33 repeat: matches 46..105 of consensus" |
| repeat_region | 35095..35842 | /note="SVA repeat: matches 3..752 of consensus" |
| repeat_region | 35853..39029 | /note="L1PAs repeat: matches 1..2855 of consensus" |
| repeat_region | 39031..39643 | /note="L1M4 repeat: matches 2135..2749 of consensus" |
| repeat_region | 39646..40249 | /note="L1Mc repeat: matches 1908..2501 of consensus" |
| repeat_region | 40250..40552 | /note="Aluub repeat: matches 1..293 of consensus" |
| repeat_region | 40553..42188 | /note="L1Mc repeat: matches 789..1908 of consensus" |
| repeat_region | 42508..42799 | /note="Aluub repeat: matches 1..299 of consensus" |
| repeat_region | 42800..43029 | /note="L1M4 repeat: matches -12..235 of consensus" |
| repeat_region | 43201..43306 | /note="53 copies 2 mer tt 65% conserved" |
| repeat_region | 43778..44031 | /note="MIR repeat: matches 3..262 of consensus" |
| repeat_region | 44186..44361 | /note="MIR repeat: matches 47..244 of consensus" |
| repeat_region | 44468..44799 | /note="MIR1A1 repeat: matches 1..364 of consensus" |
| repeat_region | 44930..45100 | /note="MER33 repeat: matches 1..187 of consensus" |
| repeat_region | 45981..46564 | /note="L1M6 repeat: matches 5573..6172 of consensus" |
| repeat_region | 46605..46734 | /note="FLAM_A repeat: matches 4..133 of consensus" |
| repeat_region | 47030..47196 | /note="MIR repeat: matches 10..164 of consensus" |
| repeat_region | 47575..48669 | /note="L1MD1 repeat: matches 5096..6213 of consensus" |
| repeat_region | 48671..50424 | /note="L1P4 repeat: matches 4376..6146 of consensus" |
| repeat_region | 50472..50939 | /note="L1MD1 repeat: matches 4641..5093 of consensus" |
| repeat_region | 50940..51248 | /note="AluY repeat: matches 1..308 of consensus" |
| repeat_region | 51249..51339 | /note="L1MD1 repeat: matches 4549..4641 of consensus" |
| repeat_region | 52657..53183 | |

```

/note="L1M4A repeat: matches 5325..5870 of consensus"
repeat_region 53242..53938
/note="L1R8 repeat: matches 31..691 of consensus"
repeat_region 53939..54218
/note="L1Pb3 repeat: matches 5867..6143 of consensus"
repeat_region 54828..55070
/note="MIR repeat: matches 2..262 of consensus"
repeat_region 56554..57365
/note="L2 repeat: matches 1826..2661 of consensus"

Query Match 6.7%; Score 21; DB 9; Length 163283;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 acatacaaaagtcagtgtggt 132
|||||
Db 75453 ACATACAAAGTCAGTGTGTGT 75433

```

[illegible]

Center: Washington University Genome Sequencing Center
 Name: Bruce Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@watson.wustl.edu
 ----- Summary Statistics -----
 Center project name: H_NH0575601

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>)

VECTOR:

pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-560014, 200 bp overlap: the clone sequenced to the right is RP5-117919, 200 bp overlap. The actual start of this clone is at base position 80345 of RP4-560014; actual end is at base position 76855 of RP5-117919.

Location/Qualifiers

FEATURES

source

```
1..57287
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2-q22"
/clone="RP11-575G1"
/clone_1lb="RPC1-11"
1..257
/rpt_family="MER1_type"
285..309
/rpt_family="AT_rich"
2332..2753
/rpt_family="L2"
3005..3078
/rpt_family="MIR"
3223..3517
/rpt_family="Alu"
3838..3903
/rpt_family="AT_rich"
complement(<4103..>50924)
/gene="WUGSC:H_NH0575601.1"
complement(join(<4103..4261,26454..26553,30330..30380,
32481..32550,50795..>50924))
/gene="WUGSC:H_NH0575601.1"
/note="match to JH0565 (PID:g87003); H_NH0575601.1"
/codon_start=1
/product="match to calcium channel alpha-2b chain
precursor"
/protein_id="AA19234.1"
/db_xref="gi:6624059"
/translating="TIVLNLMTSALDEVKRNREDDPSLIMQVSGATGARYRA
SPWNSRTPKIDLYDVRBPWYIQGAASPKDMLIVDVSQSGITLULIRVSE
MLETISDDDFVNAVSPNSMODVSCPHIVQANRKNKVLADAVNITANGITDYKKG
FSFAFEQLN"
```

```
misc_feature
4111..4261
/note="match to EST N85205 (NID:g1260830)"
4212..4256
/note="match to EST A1434840 (NID:g4298992) t113a04.x1"
5073..5153
/rpt_family="MIR"
5343..5402
/rpt_family="L2"
5457..5581
/rpt_family="MIR"
5671..6718
/rpt_family="L2"
6719..7034
/rpt_family="Alu"
7035..7558
/rpt_family="L2"
8711..9067
/rpt_family="L2"
10066..10094
/rpt_family="AT_rich"
10095..10209
/rpt_family="Alu"
10532..10555
/rpt_family="AT_rich"
10602..10653
/rpt_family="AT_rich"
11251..11256
/note="match to EST A121707 (NID:g1679348) zn80a01.r1"
11733..11927
/rpt_family="MIR"
13680..13997
/rpt_family="L2"
15639..15782
/rpt_family="MER1_type"
16263..16553
/rpt_family="Retroviral"
16364..16723
/note="match to EST A1223011 (NID:g3805214) g952e11.x1"
16810..16953
/rpt_family="MIR"
18853..18988
/rpt_family="MIR"
19017..19295
/rpt_family="Alu"
19457..19498
/rpt_family="AT_rich"
20431..20674
/rpt_family="Retroviral"
21005..21318
/rpt_family="Retroviral"
21406..21941
/rpt_family="L1"
21942..21974
/rpt_family="TA)n"
22036..22057
/rpt_family="(A)n"
22294..23118
/rpt_family="Retroviral"
23152..23172
/rpt_family="AT_rich"
23205..23294
/rpt_family="MER1_type"
24112..24354
/rpt_family="Retroviral"
25734..26165
/rpt_family="L1"
26167..26270
/rpt_family="(TA)n"
26450..26553
/note="match to EST A1434840 (NID:g4298992) t113a04.x1"
26454..26539
/note="match to EST N85205 (NID:g1260830)"
26648..26675
repeat_region
```



```

/note="contains similarity to non-LTR retroelement reverse
transcriptase
gene_id:K17E7.9"
/codon_start=1
/pseudo
/feature=not_experimental
join(36811..37599,37734..38069,38621..38818)
/note="gene_id:K17E7.10"
/codon_start=1
/evidence=not_experimental
/product="mitochondrial protein-like; contains similarity
to AAA-type ATPase"
/protein_id="BAB02995.1"
/db_xref="GI:13477082"
/translation="MMLKLIKSMVQATIRPIQYLIISYLRVYGPSPSLTHNOYTV
IEETSENGRINIVHGATQAYLEKINLDFEEREDDIYGCAXIKMRFFVKNNGNI
KPCQELREDEKRDLPVDSYIPPEYSEKAKIKSKRTLEMHITTHSHCCDWETILDH
HSSEFTYMKEDKRLITDDIDPLISKEDFTKRGRRMRYLLHGLPGAGKTSLVLA
TAKYINPDVYNTQCVKTDPTDPTRLIRVEDSILVIEDIDTSLSEGSVALSQLSL
TPWNSNGARAVITFTNKKERFDQTLICRMEMKTYMGHCEDEKTLASNYLGISHN
DAPRLYPDIRLIDGOAVTPGOVVEELMSQDVVALQSLVRYSSSSKENDHIDDL
QIPETRNNSMLDSKPKRSQTIIVEALNVCNENETEGSNIDYSSNK"
/complement(43355..43976)
/note="gb|AAFI8539.1
gene_id:K17E7.11
similar to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
join(44927..44984,45253..45347,45463..45919,46003..46159,
46227..46504,46590..46873)
/note="gene_id:K17E7.12"
/codon_start=1
/evidence=not_experimental
/product="phyllochelatin synthetase-like protein"
/protein_id="BAB02996.1"
/db_xref="GI:13477083"
/translation="MNTLPSRPSFLILFCSWNLSTEARVADLPYGNITIMDIMS
ITGDDGYAVVITFRFOYRITHEAPRGWOLGSMKKKEVIMSWGQATLEDQDSKFGN
IPHCCKTPAIVIDLPGTPYNYQIISNCRGVIISAMADPAIISFQISGSGSTN
ITVAPRNITILKAPGPGTGGPAKLVKPSRFISADKRRKRTILTNITCTYSGFLAR
KTPCCVSLSAFNETIVPCPTSCGCGNSOAGTCVDCDPSKILVAVPAVGNKNEP
LLOCTOMCPIRIVHMYKTSYKEWYKVAITNPYNNMYSOANILVYOHFNPNLTKL
FSFNKPLNPLYLINIDPAMMGIKFKNDFLSQAGPYNQSESLFKRNPDEFFEEKW
APPRRIYENGNDVMPDPDSYPMUPNASPIATSPVILLITFLSVLLIM"
50023..51531
/note="gb|AAF26945.1
gene_id:K17E7.13
similar to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
join(53764..54696,55011..55204,55298..55562)
/note="emb|CAB71883.1
gene_id:K17E7.14
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02997.1"
/db_xref="GI:13477084"
/translation="MDSVQNRDISLDPVYLVMILSFLEKDYKTSILSKRMNIG
YEAKNISKESEYVDISVDYKRSVFRVYLMNVRVPIVISFELICGPGPGE
VEIKSLIFSIROVKIYLIDFSSPMMNMGILNDDPVELPALITGLTLESITL
YACMDFSRFTVWGKILISIGWRLEKTESILSKRPLLESISICYDEVKLAQIR
VRELIIDNCITPTMYCLINPNIETIFKISGNTVDFEKVNMILLKRYEYDLFEIEN
DEPTYSKREAGDITSHLINDLSRFLTVCSYLIEVISECNDPVMLNDVQAOHLVE
TLMHPNEFTGMRLILDHCPNLETLFOLICKPEPVRLQYIDSHITLNTENISQIR
TKLILVGFCSNNEFYILNLYLPEHGFALREVELYLPPLQEMPRQMAVHGAAMLQ

```

```

QY 176 aaatttgaagaatagattgt 195
Db 42912 AAATTTGAAATATGATGCT 42893

RESULT 9
AC105599
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

1 (bases 1 to 87966)
Mazny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarta,J.,
Beaton,D., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dachorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Hulky,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karason,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kravovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisedge,H., Losado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S.,
Ogunb,M., Okunodu,G., Otagunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojokokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellod,B., Thomas,K.,
Thomas,S., Usmanu,K., Vasquez,L., Vera,V., Villalob,B., Vinson,R.,
Walli,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 87966)
Worley,K.C.
Direct Submission
Submitted (09-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNHA

```

Query Match 6.3%; Score 20; DB 8; Length 8180;
 Best Local Similarity 100.0%; Pred. No. 7.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Center clone name: CH230-127D2
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findphraplist
Consensus quality: 64078 bases at least Q40
Consensus quality: 70281 bases at least Q30
Consensus quality: 75712 bases at least Q20
Estimated insert size: 54379; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 52 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3355: contig of 3355 bp in length
3356
3455: gap of unknown length
3456
6405: contig of 2950 bp in length
6406
6505: gap of unknown length
6506
8439: contig of 1934 bp in length
8440
8539: gap of unknown length
8540
11712: contig of 3173 bp in length
11713
11812: gap of unknown length
11813
14902: contig of 3090 bp in length
14903
15002: gap of unknown length
15003
16825: contig of 1823 bp in length
16826
16925: gap of unknown length
16926
18687: contig of 1762 bp in length
18688
18787: gap of unknown length
18788
20481: contig of 1694 bp in length
20482
20581: gap of unknown length
22234: contig of 1653 bp in length
22334: gap of unknown length
22335
24360: contig of 2026 bp in length
24361
24460: gap of unknown length
24461
27456: contig of 2996 bp in length
27457
27556: gap of unknown length
29429: contig of 1873 bp in length
29529: gap of unknown length
29530
31544: contig of 2015 bp in length
31545
31644: gap of unknown length
31645
34443: contig of 2799 bp in length
34444
34543: gap of unknown length
34544
36482: contig of 1939 bp in length
36483
36582: gap of unknown length
37738: contig of 1156 bp in length
37739
37838: gap of unknown length
39324: contig of 1486 bp in length
39325
39424: gap of unknown length
40701: contig of 1277 bp in length
40801: gap of unknown length
40802
42655: contig of 1764 bp in length
42656
42665: gap of unknown length
44320: contig of 1655 bp in length
44321
44420: gap of unknown length
45953: contig of 1533 bp in length
45954
46053: gap of unknown length
47550: contig of 1497 bp in length
47551
47650: gap of unknown length
49656: contig of 2006 bp in length
49657
49756: gap of unknown length
50789: contig of 1033 bp in length
50889: gap of unknown length
50990
52357: contig of 1468 bp in length
52457: gap of unknown length
53677: contig of 1220 bp in length
53777: gap of unknown length
53678

```

```

53778
55154: contig of 1377 bp in length
55155
55254: gap of unknown length
55255
56453: contig of 1199 bp in length
56454
56553: gap of unknown length
56554
57860: contig of 1307 bp in length
57861
57960: gap of unknown length
59018: contig of 1058 bp in length
59019
59118: gap of unknown length
59119
60405: contig of 1287 bp in length
60406
60505: gap of unknown length
60506
61997: contig of 1492 bp in length
61998
62097: gap of unknown length
63125: contig of 1028 bp in length
63126
63325: gap of unknown length
64318: contig of 1093 bp in length
64319
64418: gap of unknown length
65848: contig of 1430 bp in length
65849
65948: gap of unknown length
67048: contig of 1100 bp in length
67049
67148: gap of unknown length
68348: contig of 1200 bp in length
68349
68448: gap of unknown length
68449
69891: contig of 1443 bp in length
69892
69991: gap of unknown length
71020: contig of 1029 bp in length
71021
71120: gap of unknown length
72149: contig of 1029 bp in length
72150
72249: gap of unknown length
73410: contig of 1161 bp in length
73411
74552: contig of 1042 bp in length
74553
74652: gap of unknown length
74653
75776: contig of 1024 bp in length
75777
75776: gap of unknown length
77318: contig of 1342 bp in length
77319
77418: gap of unknown length
78720: contig of 1302 bp in length
78721
78820: gap of unknown length
79845: contig of 1025 bp in length
79846
79945: gap of unknown length
79946
81305: contig of 1360 bp in length
81306
81405: gap of unknown length
82699: contig of 1294 bp in length
82700
82799: gap of unknown length
84018: contig of 1219 bp in length
84019
84118: gap of unknown length
84119
85188: contig of 1070 bp in length
85189
85288: gap of unknown length
85289
86830: contig of 1542 bp in length
86831
86930: gap of unknown length
86931
87965: contig of 1036 bp in length.
Location/Qualifiers
FEATURES
source
1..87966
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-127D2"
BASE COUNT 21580 a 20480 c 19591 g 21063 t 5252 others
ORIGIN
Query Match 6.3%; Score 20; DB 2; Length 87966;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 aagaaataactttgagaaa 40
|||||
DB 1669 AAGAAATACTTTTGAGAAA 1688
RESULT 10 AC027033 98017 bp DNA linear PLN 11-OCT-2000
AC027033/c LOCUS
DEFINITION Arabidopsis thaliana chromosome 1 BAC F21N10 genomic sequence,

```

complete sequence.
AC027033 GI:10092252
VERSION AC027033.3
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 98017)
Lin.X., Kaul.S., Town.C.D., Benito.M.-I., Creasy.T.H., Haas.B.J., Barnstead.M.E., Bowman.C.L., White.O., Nierman.W.C. and Fraser.C.M. Arabidopsis thaliana chromosome 1 BAC F21N10 genomic sequence
Unpublished
2 (bases 1 to 98017)
Lin.X. and Kaul.S.
Direct Submission
Submitted (26-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 98017)
Town.C.D. and Kaul.S.
Direct Submission
Submitted (12-SEP-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Sep 12, 2000 this sequence version replaced gi:7709919.
Address all correspondence to:atetigr.org

COMMENT
BAC clone F21N10 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SPE to T7 end of the BAC clone.
Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-061.mt.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), Glimmer4 (a variant of GlimmerM, see Mihaela Perlea, <http://www.tigr.org/softlab/glimmer.htm>), and GeneSplicer (Mihaela Perlea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/cdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).
location/Qualifiers
1. 98017

FEATURES
source
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F21N10"
184..5698
/gene="F21N10.12"
/note="pseudogene, polyprotein"
/pseudo
1368..5694
/rpt_family="ATCOP1A16|ATCOP1A16 Internal region of ATCOP1A16 copia-like LTR-retrotransposon."
1397..3853
/rpt_family="ATCOP1A11|ATCOP1A11 Internal region of ATCOP1A11 LTR-retrotransposon."
1407..1704
/rpt_family="RE1|RE1 Arabidopsis thaliana retroelement - a consensus."
3037..3529
/rpt_family="ATCOP1A11|ATCOP1A11 Internal region of

repeat_region ATCOP1A11 LTR-retrotransposon."
3037..3795
/rpt_family="ATCOP1A15|ATCOP1A15 Internal region of ATCOP1A15 copia-like LTR-retrotransposon."
3124..3348
/rpt_family="ATCOP1A17|ATCOP1A17 Internal region of ATCOP1A17 copia-like LTR-retrotransposon."
3407..3585
/rpt_family="ATCOP1A24|ATCOP1A24 Internal region of ATCOP1A24 copia-like LTR-retrotransposon."
3628..3797
/rpt_family="ATCOP1A11|ATCOP1A11 Internal region of ATCOP1A11 LTR-retrotransposon."
4256..4543
/rpt_family="ATCOP1A15|ATCOP1A15 Internal region of ATCOP1A15 copia-like LTR-retrotransposon."
4437..4629
/rpt_family="ATCOP1A11|ATCOP1A11 Internal region of ATCOP1A11 LTR-retrotransposon."
4662..4969
/rpt_family="ATCOP1A11|ATCOP1A11 Internal region of ATCOP1A11 LTR-retrotransposon."
7315..7416
/rpt_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."
7315..7416
/rpt_family="ARNOLD4|ARNOLD4 Autonomous DNA-transposon ARNOLD4."
7328..7416
/rpt_family="ARNOLD2|ARNOLD2 Autonomous DNA-transposon ARNOLD2."
7328..7416
/rpt_family="ARNOLD1|ARNOLD1 Autonomous DNA-transposon ARNOLD1."
8053..10391
/rpt_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."
8053..10391
/rpt_family="ARNOLD4|ARNOLD4 Autonomous DNA-transposon ARNOLD4."
8071..10391
/rpt_family="ARNOLD1|ARNOLD1 Autonomous DNA-transposon ARNOLD1."
8785..8913
/rpt_family="ARNOLD4|ARNOLD4 Autonomous DNA-transposon ARNOLD4."
8785..8913
/rpt_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."
8803..8913
/rpt_family="ARNOLD1|ARNOLD1 Autonomous DNA-transposon ARNOLD1."
8803..8913
/rpt_family="ARNOLD2|ARNOLD2 Autonomous DNA-transposon ARNOLD2."
8830..8913
/rpt_family="ARNOLDY2|ARNOLDY2 Non-autonomous DNA transposon ARNOLDY2 - a consensus sequence."
9506..9634
/rpt_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."
9506..9634
/rpt_family="ARNOLD4|ARNOLD4 Autonomous DNA-transposon ARNOLD4."
9525..9634
/rpt_family="ARNOLD1|ARNOLD1 Autonomous DNA-transposon ARNOLD1."
9525..9634
/rpt_family="ARNOLD2|ARNOLD2 Autonomous DNA-transposon ARNOLD2."
10984..11112
/rpt_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."

```

repeat_region      10984..11112
                    /rpt_family="ARNOLD4|ARNOLD4 Autonomously DNA-transposon
                    ARNOLD4."
repeat_region      11001..11112
                    /rpt_family="ARNOLD1|ARNOLD1 Autonomously DNA-transposon
                    ARNOLD1."
repeat_region      11001..11112
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
                    ARNOLD2."
repeat_region      complement(11870..12010)
                    /rpt_family="ATREP5|ATREP5 An AT-rich repetitive sequence
                    a consensus."
repeat_region      12462..12590
                    /rpt_family="ARNOLD4|ARNOLD4 Autonomously DNA-transposon
                    ARNOLD4."
repeat_region      12462..12590
                    /rpt_family="ARNOLD3|ARNOLD3 Autonomously DNA-transposon
                    ARNOLD3."
repeat_region      12479..12590
                    /rpt_family="ARNOLD1|ARNOLD1 Autonomously DNA-transposon
                    ARNOLD1."
repeat_region      12479..12590
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
                    ARNOLD2."
repeat_region      13255..13383
                    /rpt_family="ARNOLD4|ARNOLD4 Autonomously DNA-transposon
                    ARNOLD4."
repeat_region      13255..13383
                    /rpt_family="ARNOLD3|ARNOLD3 Autonomously DNA-transposon
                    ARNOLD3."
repeat_region      13273..13383
                    /rpt_family="ARNOLD1|ARNOLD1 Autonomously DNA-transposon
                    ARNOLD1."
repeat_region      13273..13383
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
                    ARNOLD2."
repeat_region      13975..14103
                    /rpt_family="ARNOLD3|ARNOLD3 Autonomously DNA-transposon
                    ARNOLD3."
repeat_region      13975..14103
                    /rpt_family="ARNOLD4|ARNOLD4 Autonomously DNA-transposon
                    ARNOLD4."
repeat_region      13975..14103
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
                    ARNOLD2."
repeat_region      13993..14103
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
                    ARNOLD2."
repeat_region      13993..14103
                    /rpt_family="ARNOLD1|ARNOLD1 Autonomously DNA-transposon
                    ARNOLD1."
repeat_region      15628..15837
                    /rpt_family="centromeric repeat T5H22 31150 to 31597"
                    16107..16191
                    /rpt_family="centromeric repeat T5H22 31150 to 31597"

Query Match      6.3%; Score 20; DB 8; Length 98017;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 aaatttgaataatgattgt 195
Db 29803 AAATTTGAATAATGATGTG 29784

```

```

RESULT 11
AC096828/c      AC096828      109981 bp      DNA      linear      HTG 20-DEC-2001
DEFINITION      Rattus norvegicus clone CH230-91G16, *** SEQUENCING IN PROGRESS
ACCESSION      AC096828
VERSION      AC096828.2 GI:17972957
KEYWORDS      HTG; HTGS; PHASE1.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 109981)

Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarella,J., Benton,J., Bimage,K., Blankenburg,K., Bonini,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.J., Delgado,O., Denn,A.L., Ding,Y., Dinu,H.H., Doukwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisa,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokweto,S., Oguh,M., Okunolu,G., Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N., Sisson,I., Sodergren,E., Sonalite,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R., Thomas,S., Usmani,K., Vasquez,L., Verd,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,A., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlecyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 109981)

Worley,K.C.

Direct Submission

Submitted (28-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:15799472.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGVW

Center clone name: CH230-91G16

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to findPhrapList

Consensus quality: 67454 bases at least Q40

Consensus quality: 77134 bases at least Q30

Consensus quality: 85466 bases at least Q20

Estimated insert size: 72753; sum-of-ctrls estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 0.7x in Q20 bases; sum-of-ctrls estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 59 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

1 2162: contig of 2162 bp in length
 2163 2262: gap of unknown length
 2263 4905: contig of 2643 bp in length
 4906 5005: gap of unknown length
 5006 7945: contig of 2940 bp in length
 7946 8045: gap of unknown length
 8046 9668: contig of 1623 bp in length
 9669 9769: gap of unknown length
 9769 12307: contig of 2539 bp in length
 12308 12407: gap of unknown length
 12408 14880: contig of 2473 bp in length
 14881 14980: gap of unknown length
 14981 18246: contig of 3266 bp in length
 18247 18346: gap of unknown length
 18347 20802: contig of 2456 bp in length
 20803 20902: gap of unknown length
 20903 23429: contig of 2527 bp in length
 23430 23529: gap of unknown length
 23530 26244: contig of 2715 bp in length
 26245 26344: gap of unknown length
 26345 28233: contig of 1889 bp in length
 28234 28333: gap of unknown length
 28334 30557: contig of 2224 bp in length
 30558 32242: contig of 1585 bp in length
 32243 34175: gap of unknown length
 34176 34275: gap of unknown length
 34276 37250: contig of 2974 bp in length
 37250 37349: gap of unknown length
 37350 39006: contig of 1657 bp in length
 39007 39106: gap of unknown length
 39107 40848: contig of 1742 bp in length
 40849 40948: gap of unknown length
 40949 42684: contig of 1736 bp in length
 42685 42784: gap of unknown length
 42785 45207: contig of 2423 bp in length
 45208 45307: gap of unknown length
 45308 46998: contig of 1691 bp in length
 46999 47098: gap of unknown length
 47099 48373: contig of 1275 bp in length
 48374 48473: gap of unknown length
 48474 49758: contig of 1285 bp in length
 49759 49858: gap of unknown length
 49859 51552: contig of 1694 bp in length
 51563 51652: gap of unknown length
 51653 53339: contig of 1587 bp in length
 53340 53440: gap of unknown length
 53441 54649: contig of 1310 bp in length
 54650 54749: gap of unknown length
 54750 56982: contig of 2233 bp in length
 56983 57082: gap of unknown length
 57083 58246: contig of 1164 bp in length
 58247 58346: gap of unknown length
 58347 60479: contig of 2133 bp in length
 60480 60579: gap of unknown length
 60580 62706: contig of 2127 bp in length
 62707 62806: gap of unknown length
 62807 64618: contig of 1812 bp in length
 64619 64718: gap of unknown length
 64719 66415: contig of 1697 bp in length
 66416 66515: gap of unknown length
 66516 67617: contig of 1102 bp in length
 67618 69472: gap of unknown length
 69473 69572: gap of unknown length
 69573 70847: contig of 1275 bp in length

* 70848 70947: gap of unknown length
 * 70948 72752: contig of 1805 bp in length
 * 72753 72852: gap of unknown length
 * 72853 74624: contig of 1772 bp in length
 * 74625 74725: gap of unknown length
 * 74725 76802: contig of 2078 bp in length
 * 76803 76902: gap of unknown length
 * 76903 78528: contig of 1626 bp in length
 * 78529 78628: gap of unknown length
 * 78629 80110: contig of 1482 bp in length
 * 80111 80210: gap of unknown length
 * 80211 82357: contig of 2147 bp in length
 * 82358 82457: gap of unknown length
 * 82458 83925: contig of 1468 bp in length
 * 83926 84025: gap of unknown length
 * 84026 85056: contig of 1031 bp in length
 * 85057 85156: gap of unknown length
 * 85157 86839: contig of 1683 bp in length
 * 86840 86939: gap of unknown length
 * 86940 88378: contig of 1439 bp in length
 * 88379 88478: gap of unknown length
 * 88479 89569: contig of 1091 bp in length
 * 89570 89669: gap of unknown length
 * 89670 91105: contig of 1436 bp in length
 * 91106 91205: gap of unknown length
 * 91206 92755: contig of 1550 bp in length
 * 92756 92856: gap of unknown length
 * 92857 94275: gap of unknown length
 * 94276 94375: contig of 1420 bp in length
 * 94376 96089: contig of 1714 bp in length
 * 96090 96189: gap of unknown length
 * 96190 97291: contig of 1102 bp in length
 * 97292 97391: gap of unknown length
 * 97392 98457: contig of 1066 bp in length
 * 98458 98557: gap of unknown length
 * 98559 99593: contig of 1036 bp in length
 * 99594 99693: gap of unknown length
 * 99694 100830: contig of 1137 bp in length
 * 100831 100930: gap of unknown length
 * 100931 102463: contig of 1533 bp in length
 * 102464 102563: gap of unknown length
 * 102564 104121: contig of 1558 bp in length
 * 104122 104221: gap of unknown length
 * 104222 105573: contig of 1352 bp in length
 * 105574 105673: gap of unknown length

Query Match 5.3% Score 20: DB 2: Length 109981;
 Best Local Similarity 100.0% Pred. No. 7;
 Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

OY 44 aacaaatgaatcagttct 63
 DB 93606 AACAAATGAATCAGTTCT 93587

RESULT 12
 LOCUS AC008271 111159 bp DNA linear PRI 09-MAY-2001
 DEFINITION Homo sapiens BAC clone RP11-123E16 from 2, complete sequence.
 AC008271
 VERSION AC008271.6 GI:11120938
 KEYWORDS HTG.

SOURCE

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

2 (bases 1 to 111159)
 Haakenson, B., Stoneking, T., Elliott, G. and Phillips, A.


```

* 18586 18685: gap of unknown length
* 18686 22719: contig of 4034 bp in length
* 22720 22819: gap of unknown length
* 22820 25644: contig of 2825 bp in length
* 25645 25744: gap of unknown length
* 25745 27426: contig of 1682 bp in length
* 27427 27526: gap of unknown length
* 27527 28729: contig of 1203 bp in length
* 28730 28829: gap of unknown length
* 28830 30452: contig of 1623 bp in length
* 30453 30552: gap of unknown length
* 30553 33530: contig of 2978 bp in length
* 33531 33630: gap of unknown length
* 33631 36409: contig of 2779 bp in length
* 36410 36509: gap of unknown length
* 36510 37569: contig of 1060 bp in length
* 37570 37669: gap of unknown length
* 37670 38852: contig of 1183 bp in length
* 38853 38952: gap of unknown length
* 38953 40293: contig of 1341 bp in length
* 40294 40393: gap of unknown length
* 40394 42174: contig of 1781 bp in length
* 42175 42274: gap of unknown length
* 42275 44226: contig of 1952 bp in length
* 44227 44326: gap of unknown length
* 44327 45602: contig of 1276 bp in length
* 45603 45702: gap of unknown length
* 45703 46336: contig of 1234 bp in length
* 46337 47036: gap of unknown length
* 47037 49538: contig of 2502 bp in length
* 49539 51322: contig of 1684 bp in length
* 51323 51422: gap of unknown length
* 51423 54203: contig of 2781 bp in length
* 54204 54503: gap of unknown length
* 54504 56244: contig of 1941 bp in length
* 56245 56344: gap of unknown length
* 56345 58676: contig of 2332 bp in length
* 58677 58776: gap of unknown length
* 58777 60187: contig of 1411 bp in length
* 60188 60287: gap of unknown length
* 60288 62280: contig of 1933 bp in length
* 62281 64324: gap of unknown length
* 64325 64424: gap of unknown length
* 64425 66005: contig of 1581 bp in length
* 66006 66105: gap of unknown length
* 66106 68093: contig of 1988 bp in length
* 68094 68193: gap of unknown length
* 68194 69970: contig of 1777 bp in length
* 69971 70070: gap of unknown length
* 70071 71599: contig of 1529 bp in length
* 71600 71699: gap of unknown length
* 71700 72859: contig of 1160 bp in length
* 72860 72959: gap of unknown length
* 72960 74333: contig of 1374 bp in length
* 74334 74433: gap of unknown length
* 74434 76033: contig of 1600 bp in length
* 76034 76133: gap of unknown length
* 76134 78122: contig of 1989 bp in length
* 78123 80115: contig of 1793 bp in length
* 80116 80115: gap of unknown length
* 80116 81418: contig of 1303 bp in length
* 81419 81518: gap of unknown length
* 81519 83009: contig of 1491 bp in length
* 83010 83109: gap of unknown length
* 83110 84715: contig of 1606 bp in length
* 84716 84815: gap of unknown length
* 84816 86128: contig of 1313 bp in length
* 86129 86228: gap of unknown length
* 86229 88570: contig of 2342 bp in length
* 88571 88670: gap of unknown length

```

```

* 88671 89768: contig of 1098 bp in length
* 89769 89868: gap of unknown length
* 89869 90939: contig of 1071 bp in length
* 90940 91039: gap of unknown length
* 91040 92547: contig of 1508 bp in length
* 92548 92647: gap of unknown length
* 92648 94041: contig of 1394 bp in length
* 94042 94141: gap of unknown length
* 94142 95517: contig of 1376 bp in length
* 95518 95617: gap of unknown length
* 95618 96690: contig of 1073 bp in length
* 96691 96790: gap of unknown length
* 96791 97926: contig of 1136 bp in length
* 97927 98026: gap of unknown length
* 98027 99504: contig of 1478 bp in length
* 99505 99604: gap of unknown length
* 99605 101254: contig of 1650 bp in length
* 101255 101354: gap of unknown length
* 101355 103141: contig of 1787 bp in length
* 103142 103241: gap of unknown length
* 103242 104438: contig of 1197 bp in length
* 104439 104538: gap of unknown length
* 104539 105718: contig of 1180 bp in length
* 105719 105818: gap of unknown length
* 105819 106835: contig of 1017 bp in length
* 106836 106935: gap of unknown length
* 106936 108585: contig of 1650 bp in length
* 108586 108685: gap of unknown length
* 108686 109842: contig of 1157 bp in length
* 109843 109942: gap of unknown length
* 109943 111491: contig of 1549 bp in length

Query Match      6.3%; Score 20; DB 2; Length 114927;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21 aagaataactcttgagaa 40
Db 96146 AAGAAATACCTTTGAGAA 96165
|||||
AC094962
LOCUS      14
DEFINITION Rattus norvegicus clone CH230-6L19, *** SEQUENCING IN PROGRESS ***,
AC094962
ACCESSION AC094962
VERSION   AC094962.2 GI:17941763
KEYWORDS  HTG; HTGS PHASE1.
SOURCE    Norway rat.
ORGANISM  Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE  1 (bases 1 to 115948)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Alf-oman,F.R., Allen,C.,
           Albrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
           Benton,J., Bimarge,K., Blankenburg,K., Bonin,D., Bouck,J.,
           Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
           Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
           Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R.,
           Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
           Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
           Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
           Demn,A.L., Ding,Y., Dinh,H.H., Doultwaite,K.J., Draper,H.,
           Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
           Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
           Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
           Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
           Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
           Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
           Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
           Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,

```

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nockekko,S., Ogih,M., Okunou,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 115948)
Worley,K.C.

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624799.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GRUX
Center clone name: CH230-6L19
----- Summary Statistics
findhraplust
Assembly program: Phrap; version 0.990329first call to
Consensus quality: 81103 bases at least Q40
Consensus quality: 94888 bases at least Q30
Consensus quality: 104411 bases at least Q20
Estimated insert size: 51502; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-1p estimation
Quality coverage: 0.5x in Q20 bases; sum-of-coverage estimation

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
* consists of 71 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1010: contig of 1010 bp in length
* 1011 1110: gap of unknown length
* 1111 4665: contig of 3555 bp in length
* 4666 4765: gap of unknown length
* 4766 8651: contig of 3886 bp in length
* 8652 8752 11403: gap of unknown length
* 8752 11403: contig of 2652 bp in length
* 11404 11503: gap of unknown length
* 11504 13360: contig of 1857 bp in length
* 13361 13460: gap of unknown length
* 13461 15610: contig of 2150 bp in length
* 15611 15710: gap of unknown length
* 15711 18255: contig of 2545 bp in length
* 18256 18355: gap of unknown length
* 18356 19947: contig of 1592 bp in length

19948 20047: gap of unknown length
* 20048 21200: contig of 1153 bp in length
* 21201 21300: gap of unknown length
* 21301 23457: contig of 2157 bp in length
* 23458 23557: gap of unknown length
* 23558 26085: contig of 2528 bp in length
* 26086 26185: gap of unknown length
* 26186 27472: contig of 1287 bp in length
* 27473 27572: gap of unknown length
* 27573 29081: contig of 1509 bp in length
* 29082 29181: gap of unknown length
* 29182 30555: contig of 1374 bp in length
* 30556 30655: gap of unknown length
* 30656 32892: contig of 2237 bp in length
* 32893 32992: gap of unknown length
* 32993 34484: contig of 1492 bp in length
* 34485 34584: gap of unknown length
* 34585 36413: contig of 1829 bp in length
* 36414 36513: gap of unknown length
* 36514 38275: contig of 1762 bp in length
* 38276 38375: gap of unknown length
* 38376 39975: contig of 1600 bp in length
* 39976 40076: gap of unknown length
* 40076 41487: contig of 1412 bp in length
* 41488 41587: gap of unknown length
* 41588 42910: contig of 1323 bp in length
* 42911 43010: gap of unknown length
* 43011 44419: contig of 1409 bp in length
* 44420 44519: gap of unknown length
* 44520 45673: contig of 1154 bp in length
* 45674 45773: gap of unknown length
* 45774 47130: contig of 1357 bp in length
* 47131 47230: gap of unknown length
* 47231 48523: contig of 1633 bp in length
* 48524 49024: gap of unknown length
* 49024 50794: contig of 1771 bp in length
* 50795 50894: gap of unknown length
* 50895 52255: contig of 1361 bp in length
* 52256 52355: gap of unknown length
* 52356 53507: contig of 1152 bp in length
* 53508 53607: gap of unknown length
* 53608 54836: contig of 1229 bp in length
* 54837 54936: gap of unknown length
* 54937 56428: contig of 1492 bp in length
* 56429 56528: gap of unknown length
* 56529 57735: contig of 1207 bp in length
* 57736 57835: gap of unknown length
* 57836 58872: contig of 1037 bp in length
* 58873 58972: gap of unknown length
* 58973 60518: contig of 1546 bp in length
* 60519 60618: gap of unknown length
* 60619 61655: contig of 1037 bp in length
* 61656 61755: gap of unknown length
* 61756 63254: contig of 1999 bp in length
* 63255 63354: gap of unknown length
* 63355 64932: contig of 1578 bp in length
* 64933 65032: gap of unknown length
* 65033 66962: contig of 1930 bp in length
* 66963 67062: gap of unknown length
* 67063 68267: contig of 1205 bp in length
* 68268 68367: gap of unknown length
* 68368 69887: contig of 1520 bp in length
* 69888 71612: gap of unknown length
* 71613 71712: contig of 1625 bp in length
* 71713 73377: gap of unknown length
* 73378 73477: contig of 1665 bp in length
* 73478 75473: gap of unknown length
* 75474 75573: contig of 1996 bp in length
* 75574 76953: gap of unknown length
* 76954 77053: contig of 1380 bp in length
* 77054 78130: gap of unknown length
* 78131 78230: gap of unknown length

* 78231 79584: contig of 1354 bp in length
* 79585 79684: gap of unknown length
* 79685 80811: contig of 1127 bp in length
* 80812 80911: gap of unknown length
* 80912 82052: contig of 1141 bp in length
* 82053 82152: gap of unknown length
* 82153 83643: contig of 1491 bp in length
* 83644 83743: gap of unknown length
* 83744 85227: contig of 1484 bp in length
* 85228 85328: gap of unknown length
* 85328 86706: contig of 1379 bp in length
* 86707 86806: gap of unknown length
* 86807 87892: contig of 1086 bp in length
* 87893 87992: gap of unknown length
* 87993 89432: contig of 1440 bp in length
* 89433 89532: gap of unknown length
* 89533 90625: contig of 1093 bp in length
* 90626 90725: gap of unknown length
* 90726 91896: contig of 1171 bp in length
* 91897 91996: gap of unknown length
* 91997 93048: contig of 1052 bp in length
* 93049 93148: gap of unknown length
* 93149 94569: contig of 1421 bp in length
* 94570 94669: gap of unknown length

Query Match 6.3%; Score 20; DB 2; Length 115948;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 acagagattacttctgca 277
|||||
Db 70157 ACAGAGATCATTTACTTGCA 70176

RESULT 15
HSJ100419/c 117757 bp DNA linear PRI 15-MAR-2001
LOCUS
DEFINITION
Human DNA sequence from clone RP5-100419 on chromosome 20 Contains
ESTs, STSS and GSSs, complete sequence.
ACCESSION
AL121912
VERSION
AL121912.18 GI:7018383
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 117757)
AUTHORS
Wallis, J.
TITLE
Direct Submission
JOURNAL
Submitted (09-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerysanger.ac.uk
COMMENT
On Feb 21, 2000 this sequence version replaced gi:6983194.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
This sequence is the entire insert of clone RP5-100419 The true
left end of clone RP11-500011 is at 24031 in this sequence. This
sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

FEATURES
source

30): an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP5-100419 is from
the library RPc1-5 constructed by the group of Pieter de Jong. For
further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.
location/Qualifiers
1..117757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone_1lb="RPc1-5"
265..312
/note="24 copies 2 mer aa 728 conserved"
562..692
/note="MIR repeat: matches 51..195 of consensus"
826..913
/note="MIR repeat: matches 272..354 of consensus"
914..1280
/note="THE1B repeat: matches 1..364 of consensus"
1291..1461
/note="THE1B repeat: matches 1..364 of consensus"
2025..2173
/note="THE1B repeat: matches 81..272 of consensus"
2213..2294
/note="MIR repeat: matches 2344..2492 of consensus"
2295..2422
/note="MIR repeat: matches 85..169 of consensus"
2439..2808
/note="MIR repeat: matches 1..128 of consensus"
2913..2975
/note="THE1B repeat: matches 1..364 of consensus"
4751..5056
/note="MIR repeat: matches 145..207 of consensus"
5841..6068
/note="MIR repeat: matches 2..296 of consensus"
6078..6129
/note="MIR repeat: matches 1..236 of consensus"
6133..6977
/note="26 copies 2 mer tt 848 conserved"
6978..7495
/note="LIP43 repeat: matches 5286..6146 of consensus"
7506..8207
/note="LIP43 repeat: matches 4747..5264 of consensus"
complement(8496..8966)
/note="LIP43 repeat: matches 5476..6180 of consensus"
8936..9155
/note="match: GSS: Em:AQ487745"
9122..9174
/note="MIR repeat: matches 8..241 of consensus"
9898..9919
/note="L2 repeat: matches 2668..2735 of consensus"
complement(10194..11028)
/note="11 copies 2 mer at 100% conserved"
10451..10635
/note="match: GSS: Em:AQ738844"
10451..10635
/note="MIR repeat: matches 2..184 of consensus"
11814..12073
/note="L2 repeat: matches 2468..2750 of consensus"
12588..12796
/note="MIR repeat: matches 23..236 of consensus"
15726..16149
/note="L2 repeat: matches 5050..5473 of consensus"
16261..16628
/note="LIP43 repeat: matches 5050..5473 of consensus"
17825..18145
/note="MIR repeat: matches 1..363 of consensus"
/note="LIP43 repeat: matches 1..311 of consensus"
complement(18234..18628)
/note="match: GSS: Em:AQ549175"
18512..18622
/note="MIR repeat: matches 6..113 of consensus"

```

repeat_region 18843. .18929
/note="MER5A repeat: matches 91. .171 of consensus"
repeat_region 20014. .20350
/note="THEIR repeat: matches 1. .364 of consensus"
repeat_region 20477. .20532
/note="7 copies 8 mer cacacacc 80% conserved"
repeat_region 20490. .20531
/note="21 copies 2 mer ac 83% conserved"
repeat_region 20501. .20532
/note="8 copies 4 mer caca 93% conserved"
repeat_region 21385. .21805
/note="LMC4 repeat: matches 7288. .7768 of consensus"
repeat_region 22623. .22714
/note="MIR repeat: matches 39. .131 of consensus"
repeat_region 22800. .22931
/note="LIME3A repeat: matches 6015. .6157 of consensus"
repeat_region 23151. .23198
/note="24 copies 2 mer ga 72% conserved"
repeat_region 23918. .23978
/note="MIR repeat: matches 28. .90 of consensus"
repeat_region 24098. .24217
/note="15 copies 8 mer ttctctt 63% conserved"
repeat_region 24152. .24215
/note="16 copies 4 mer tctt 68% conserved"
repeat_region 24365. .24469
/note="MER5B repeat: matches 66. .178 of consensus"
repeat_region 25225. .25342
/note="Alusx repeat: matches 1. .307 of consensus"
repeat_region 25746. .26044
/note="Aluub repeat: matches 1. .307 of consensus"
repeat_region 26883. .26885
/note="clone RP11-500011
gac in this entry
substitution"
repeat_region 26988. .27196
/note="LMC4 repeat: matches 6626. .6848 of consensus"
repeat_region 27203. .27350
/note="MIRB repeat: matches 237. .389 of consensus"
repeat_region 27478. .27537
/note="10 copies 8 mer aaagagag 67% conserved"
repeat_region 27479. .27535
/note="19 copies 3 mer gag 71% conserved"
repeat_region 27672. .27674
/note="clone RP11-500011
tga in this entry
substitution"
repeat_region 28086. .28195
/note="L2 repeat: matches 2623. .2736 of consensus"
repeat_region 28412. .28608
/note="MIR repeat: matches 40. .250 of consensus"
repeat_region 29144. .29255
/note="Charlie2 repeat: matches 1. .114 of consensus"
repeat_region 29304. .29378
/note="Charlie2 repeat: matches 233. .306 of consensus"
repeat_region 29664. .29666
/note="clone RP11-500011
aac in this entry
substitution"
variation 29784. .29786
/replace="agc"
/note="clone RP11-500011
cat in this entry
substitution"
variation 29849. .29851
/replace="cgt"
/note="clone RP11-500011
acg in this entry
substitution"
repeat_region 30483. .30544
/note="31 copies 2 mer tt 72% conserved"

```

```

variation 30629. .30631
/note="clone RP11-500011
aac in this entry
substitution"
repeat_region 30668. .30795
/replace="agc"
/note="MIR repeat: matches 41. .173 of consensus"
variation 32436. .32438
/note="clone RP11-500011
caa in this entry
substitution"
variation 33016. .33018
/replace="cca"
/note="clone RP11-500011
cct in this entry
substitution"
/replace="ctt"

```

```

Query Match 6.3%; Score 20; DB 9; Length 117757;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 aaatgaatcgtttctcca 66
|||||
Db 37386 AATGAATCAGTTTCCTCCA 37367

```

Search completed: September 16, 2002, 06:31:07
 Job time: 12345 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:05:22 : Search time 3813.96 Seconds
(without alignments)
9217.868 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 ggtcgtcagagataaagtt.....tcaataaacacttgctctgtg 1680

Scoring table:

OLIGO_MNC
Gapop 60.0, Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Query Match Length DB ID Description

| | | | | | | |
|----|------|------|--------|----|-----------|--------------------|
| 1 | 1628 | 96.9 | 1680 | 6 | AX224643 | AX224643 Sequence |
| 2 | 1457 | 86.7 | 1636 | 6 | AX092346 | AX092346 Sequence |
| 3 | 1457 | 86.7 | 2036 | 6 | AX224647 | AX224647 Sequence |
| 4 | 1355 | 80.7 | 1684 | 9 | BC008429 | BC008429 Homo sapi |
| 5 | 562 | 33.5 | 1035 | 6 | AX301908 | AX301908 Sequence |
| 6 | 405 | 24.1 | 963 | 6 | AX301906 | AX301906 Sequence |
| 7 | 181 | 10.8 | 159440 | 9 | AL355392 | AL355392 Human DNA |
| 8 | 161 | 9.6 | 161593 | 9 | HSBA49G10 | AL321901 Human DNA |
| 9 | 26 | 1.5 | 26 | 6 | AX301961 | AX301961 Sequence |
| 10 | 26 | 1.5 | 89536 | 2 | AC094188 | AC094188 Rattus no |
| 11 | 24 | 1.4 | 1629 | 10 | MMU46068 | U46068 Mus musculu |
| 12 | 22 | 1.3 | 200000 | 6 | AP000494 | AP000494 Homo sapi |
| 13 | 21 | 1.3 | 21 | 6 | AX301962 | AX301962 Sequence |
| 14 | 21 | 1.2 | 667 | 10 | AF184920 | AF184920 Rattus no |
| 15 | 21 | 1.2 | 32369 | 2 | AC096278 | AC096278 Rattus no |
| 16 | 21 | 1.2 | 75382 | 2 | AC022091 | AC022091 Homo sapi |
| 17 | 21 | 1.2 | 114748 | 2 | AC097411 | AC097411 Rattus no |
| 18 | 21 | 1.2 | 150207 | 2 | AL451067 | AL451067 Homo sapi |
| 19 | 21 | 1.2 | 156012 | 2 | AC008453 | AC008453 Homo sapi |
| 20 | 21 | 1.2 | 169045 | 2 | AC092972 | AC092972 Homo sapi |
| 21 | 21 | 1.2 | 178377 | 9 | AC007685 | AC007685 Homo sapi |
| 22 | 21 | 1.2 | 187651 | 2 | AC055122 | AC055122 Homo sapi |
| 23 | 21 | 1.2 | 187727 | 2 | AC021142 | AC021142 Homo sapi |
| 24 | 21 | 1.2 | 188527 | 9 | AC084357 | AC084357 Homo sapi |
| 25 | 21 | 1.2 | 231743 | 2 | AC095186 | AC095186 Rattus no |
| 26 | 21 | 1.2 | 248087 | 2 | AC097742 | AC097742 Rattus no |
| 27 | 20 | 1.2 | 500 | 9 | IR057017 | AL110291 Homo sapi |
| 28 | 20 | 1.2 | 1321 | 1 | AF246363 | AF246363 Acidithio |
| 29 | 20 | 1.2 | 1698 | 10 | AB041660 | AB041660 Mus muscu |
| 30 | 20 | 1.2 | 1788 | 9 | HSHP512 | X12433 Human PHS1- |
| 31 | 20 | 1.2 | 1961 | 9 | AB060916 | AB060916 Macaca fa |
| 32 | 20 | 1.2 | 8032 | 6 | AX069334 | AX069334 Sequence |
| 33 | 20 | 1.2 | 12827 | 2 | AC107283 | AC107283 Rattus no |
| 34 | 20 | 1.2 | 20058 | 8 | AC023166 | AC023166 Mus muscu |
| 35 | 20 | 1.2 | 59617 | 8 | AC007087 | AC007087 Arabidops |
| 36 | 20 | 1.2 | 71225 | 9 | AL358196 | AL358196 Human DNA |
| 37 | 20 | 1.2 | 71476 | 2 | AP000490 | AP000490 Homo sapi |
| 38 | 20 | 1.2 | 72921 | 2 | AC102273 | AC102273 Mus muscu |
| 39 | 20 | 1.2 | 140596 | 2 | RN75P15 | AL603730 Rattus no |
| 40 | 20 | 1.2 | 144333 | 2 | AC097962 | AC097962 Rattus no |
| 41 | 20 | 1.2 | 147109 | 9 | AL591043 | AL591043 Human DNA |
| 42 | 20 | 1.2 | 149394 | 2 | AP000828 | AP000828 Homo sapi |
| 43 | 20 | 1.2 | 150000 | 2 | AC095238 | AC095238 Rattus no |
| 44 | 20 | 1.2 | 162481 | 2 | AC091864 | AC091864 Homo sapi |
| 45 | 20 | 1.2 | 171304 | 2 | AC090297 | AC090297 Homo sapi |

ALIGNMENTS

RESULT 1

AX224643

LOCUS AX224643 1680 bp DNA linear PAT 10-SEP-2001

DEFINITION Sequence 1 from Patent WO0161055.

ACCESSION AX224643

VERSION AX224643.1 GI:15554772

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1680)

AUTHORS Chen,S.Y., Sun,Y. and Macina,R.A.

TITLE Methods for diagnosing, monitoring, staging, imaging and treating lung cancer via lung cancer specific genes

JOURNAL Patent: WO 0161055-A 1 23-AUG-2001;

diabexus, Inc. (US)

FEATURES

source location/Qualifiers

1..1680

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 380 a 503 c 429 g 367 t 1 others

ORIGIN

Query Match 96.9%; Score 1628; DB 6; Length 1680;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 53 ggaagcgagcgaggaactcagctgcccaggtgctccatccctcgcctcgtcctcgtc 112
Db 53 GGAAGCGGCGGAGACTCCAGCGCTGGCCAGGCTGCGATCCCTGCTGCTGCTGCTG 112
Oy 113 caccctggaagatgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 112
Db 113 CACCCTGGAAGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 112
Oy 173 ccttgatcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 172
Db 173 CCTTGATCCAGAACCCAGCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCT 172
Oy 233 aagaaagctgacacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 232
Db 233 AAGAAAGCTGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 232
Oy 293 cgtctcagatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 292
Db 293 CGCTCTCAGTGCATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 292
Oy 353 gaaacacgctcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 352
Db 353 GAACACCGCTCTGAGACATCATCTGCTGAGGCTGATCAGGCTGATCAGGCTGATCAG 352
Oy 413 gcaagctgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 412
Db 413 GCAAGCTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 412
Oy 473 gctctgagctcagaacgagcagcagcagcagcagcagcagcagcagcagcagcagc 472
Db 473 GCTCTGAGCTCAGAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 472
Oy 533 ccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 532
Db 533 CCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 532
Oy 593 ctgtgcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 592
Db 593 CTGTGCAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 592
Oy 653 gaaacgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 652
Db 653 GAAACGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 652
Oy 713 aaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 712
Db 713 AAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 712
Oy 773 gctctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 772
Db 773 GCTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 772
Oy 833 tgcacatcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 832
Db 833 TGCCATCAAGGCTGACACCATCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCT 832
Oy 893 aaagctgacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 892
Db 893 AAAGCTGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 892
Oy 953 caccacgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 952
Db 953 CACCACGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 952
Oy 1012 ctctcagaagaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1012
Db 1012 CTCTCAGAAGAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1012
Oy 1072 ctctcagaagaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1072
Db 1072 CTCTCAGAAGAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1072

```

```

Oy 1073 gaagtcagacatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1132
Db 1073 GAAGTCAGACATGGGCTGATCATATGAAGCTGAGTAAAGCTGAGTAAAGCTGAGT 1132
Oy 1133 cgtgaagatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1192
Db 1133 CGTGAAGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1192
Oy 1193 ggcacacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1252
Db 1193 GGCACACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1252
Oy 1253 ccttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1312
Db 1253 CCTTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1312
Oy 1313 caacttgaatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1372
Db 1313 CAACCTTGAAATACATCAGCTGATCGGATCCAGCTGATCGGATCCAGCTGATCG 1372
Oy 1373 ccaacatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1432
Db 1373 CCAACATGAGCTGTAAGAAACATCATCATCATCATCATCATCATCATCATCATCAT 1432
Oy 1433 ccaagatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1492
Db 1433 CCAAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1492
Oy 1493 ggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1552
Db 1493 GGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1552
Oy 1553 acccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1612
Db 1553 ACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1612
Oy 1613 ccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1672
Db 1613 CCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1672
Oy 1673 tgcctgtg 1680
Db 1673 TGCTGTG 1680

```

RESULT 2
 AX092346
 LOCUS
 DEFINITION
 AX092346
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 1 (bases 1 to 1636)
 Eaton,D.L., Flivaroff,E., Gerritsen,M.E., Goddard,A.,
 Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
 Wood,W.I.
 TITLE
 Secreted and transmembrane polypeptides and nucleic acids encoding
 the same
 JOURNAL
 Patent: WO 0116318-A 77 08-MAR-2001;
 Genentech, Inc. (US)
 FEATURES
 source
 1..1636
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 374 a 492 c 411 g 359 t
 ORIGIN

QY 53 ggaagcgccgagagactccagcgtgccaggtctgcacatcctgcactgtgcctctga 112
 Db 92 GGAGCGGGCCGAGAGATCCAGCGGTGGCCAGATCCCGCACTTCCTCTGA 151
 QY 113 caactggaaagatggccggcccgctgaccttaacctctctgtgtgttgctgagcga 172
 Db 152 CACCTGGGAAGATGGCCGGCCCGGTGACCTTCACCTTCCTGTGTGTGGCTGAGCACA 211
 QY 173 ccttgatccaagccacctcaagtlcccaatgcagttctcatctctggcccaaaagatca 232
 Db 212 CTTTATCCAAAGCCACCTCACTCCACATGCAGTTCTCATCTCGGCCCAAAAGTCATCA 271
 QY 233 aagaaagctgacacaggaagctgaaagaccacacagccacagcaatccctgacagctgc 292
 Db 272 AAGAAAGCTGACACAGAGAGCTGAAGAGCACACAGCCACAGCATCCTGAGAGAGCTGC 331
 QY 293 cgtctcagtgccaatggcggaagaaagccagccgagag -atcccgctgtgtggagcctgg 351
 Db 332 CGCTGCTCAGTGCATGGGGGGAAGAACCCAGCGAGAGCATCCCTGTGTGTGGGAGCCTGG 391
 QY 352 tgaacacgcgtccctgaaacacatcatctgtgtgaaggtcatcacagactaacaatccctcagc 411
 Db 392 TGAACACGCTCTGAAGACACRTCACTGTGCTGAAGGTCAATCAGCTAACATCCTCCAGC 451
 QY 412 tgcaggtgaagccctcggccaaatgaacagagagctgtcagttcaagaatccctggaatgg 471
 Db 452 TGAGAGTAAACCTCGGCGCAATGACAGAGAGCTGTAGTCAAGATCCCTGTGACATGG 511
 QY 472 tgcgtgattcaacacgcccctgtgtcaagacacatcgtgtgaagttcacatagactgag 531
 Db 512 TGCTGTGATCAACACGCCCTCGTGTCAAGAACATCGTGAAGTTCCACATACAGCTGAGG 571
 QY 532 cccaagacacatccgcatgaggaacacagttgcaagtgcccaacccgctgtgctcaagt 591
 Db 572 CCAAGCACCATCGCATGACACACAGTCAAGTGGCCCGCCCGCTGCTGCTCAGTG 631
 QY 592 actgtgcacacagcgaatggagccctgtgcacatcccaatgcagactgtctctctcctgg 651
 Db 632 ACTGTGCACACAGCATGGAGAGCTGGGCACTCAACTGCTGCAATGAAGCTTCCTCTCG 691
 QY 652 tgaagccttaagtaagcaggtcatgaacctcctagtgccatccctgcaccaatagtaga 711
 Db 692 TGAACGCTTAAGCTAAGCAGGTGATGAACCTCTAGTGCATCCCGCCCAATCTAGTGA 751
 QY 712 aaacaacagctgtgtccgcatgagagacttctctcaatgagcatgtatgacagactctgc 771
 Db 752 AAAACAGAGCTGTCCCGTATGAGAGCTTCCTCAATGAGCATGATGAGACCTCTCTGC 811
 QY 772 agctgtgaaagtgcacatttccctcagatgacagctgtgagttgacatttgatc 831
 Db 812 AGCTGTGAGAGTGCCCATTTCCCTAGCATTGACCGGTGTGAGTTGACCTTCTGTATC 871
 QY 832 ctgcacaaaggttgaccacttccagcttacttactggtgggccaagtgtgtgactacag 891
 Db 872 CTGCCATCAAGGGTGAACCATTCACCTCTACCTGGGGCCCAAGTGTGTGACTACAGG 931
 QY 892 gaaaggtgaccaaagtgttcaataactctgcagcttccctgcacaaatgcccacccggagca 951
 Db 932 GAAAGGTGACCAAGTGTTCATTAATCTGACAGCTTCCTGACATGCCCAACCCGGAGCA 991
 QY 952 agatcccggtgagctcactcgtgaaatcagagcgtgtgtgaaagctgcagtgagctgtgc 1011
 Db 992 ACATCCCGTTCAGCTCATCTGAGTACAGAGAGTGTGAAAGCTCACTGGCTGTATC 1051
 QY 1012 tctctcagaagaattcatctgtcctgtgtgagactgtgtctctcctgagagtgccatcggc 1071
 Db 1052 TCTCTCAGAAGAATTCTATGCTGCTGTGTGACTGTGTCTCTGAGAGTCCCATCGGC 1111
 QY 1072 tgaagtcagcatcggtgtgtcaatgaaagagctgcagataagctgtgacttcaacaga 1131
 Db 1112 TGAAGTCAAGCATCGGGCTGATCATGAAGAGGTGCGAGATAAAGCTGGAGATCTTCCACA 1171
 QY 1132 tcgtgaagatcctaactaaggaactcccgagtttlttatagaccagggccaatgccaag 1191

Db 1172 TCGTGAAGATCTTAACCTACAGACACTCCAGATTTTATTATAGACCAAGGCCATGCCAAG 1231
 QY 1192 tggcccaactgatatgtcttggaagtgttccctccagtgaaagccctccgacctgttca 1251
 Db 1232 TGCGCCAACTGATGCTCTGTGAAAGTGTTCCTTCAGATGAAGCCCGCCCTTTCTTCA 1291
 QY 1252 ccttggacatcgaagccagcttcggaagctcagtttcaacacaaaggtgacaaattatc 1311
 Db 1292 CCTGGGCAATCAAGAGCAGCTCGAAGCTCAGTTTACACCAAGGTGACCAACTTATAC 1351
 QY 1312 tcaacttgaataacatcagcttcgtatcgatccagctgataactctgagttgctgtgt 1371
 Db 1352 TCAACTTGAATAACATCAGCTGTGATCGGATCCAGCTGATGAACCTGTGGATTGGCTGT 1411
 QY 1372 tccaactgtatgtctgaaacacatcatcaactgataatcaatccatccctgcgcga 1431
 Db 1412 TCCAACTGTATGTTCGAAACATCATCATGATGATCACTCCATCTCAGTGGCGGA 1471
 QY 1432 accaagatggcaaatgaatctgggtgccagtgatcattgtgtgaagccttgagttcg 1491
 Db 1472 ACCAGATGGCAAAATTAAATCTGGGGTCCAGTGTCAATGGTGAAGGCCTTGGGATTCG 1531
 QY 1492 aggaagctgagtcctcactcagcacaagagatgccttgtgttactccagctcctctgtga 1551
 Db 1532 AGGACACTGATGCTCAGTACAGTACAGAGATGCCCTTGTGCTTACTCCACCTCTTGTGA 1591
 QY 1552 aaccaagctcctcctgtctccagtgaaagcttgaatggcagccalcaggaagctgtgt 1611
 Db 1592 AACCCAGCTCTCTCTGCTCCAGTGAAGACTTGATGGCAGGCATCAAGGAAGCTGGGT 1651
 QY 1612 cccagctggagatgggtgtgagctctatagacacatccctctctgaaatcaataaac 1671
 Db 1652 CCCAGCTGGAATGAGGTGATGAGCTTATGACCATCTCTGCAATCAATAAATACAC 1711
 QY 1672 ttgcctgtg 1680
 Db 1712 TTGCTGTG 1720

RESULT 4
LOCUS BC008429 1684 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, Similar to DNA segment, Chr 2, Massachusetts Institute of Technology 19, clone MGC:14597, IMAGE:4291561, mRNA, complete cds.
ACCESSION BC008429
VERSION BC008429.1 GI:14250057
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 1684)
TITLE Strausberg, R.
JOURNAL Direct Submission
 Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLOMTECH
 cDNA Library Preparation: CLOMTECH
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) medepax11.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILMIL at: <http://image.llnl.gov>
Series: IRAL, Plate: 21, Row: 3, Column: 12
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES

source

1. 1684

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:14597 IMAGE:4291561"

/tissue_type="Skeletal Muscle"

/clone_lib="NIH_MGC_81"

/lab_host="DH10B"

/note="Vector: pDNR-LIB"

CDS

77..1531

/codon_start=1

/product="Similar to DNA segment, Chr 2, Massachusetts

Institute of Technology 19"

/protein_id="AAH08429.1"

/db_xref="GI:1425058"

/translation="MAGPWTLLCGLLAATLILQATLSPTVLLIGPVYIEKLTQEL
KDHNAATSLQQLPLISAMREKPAIGIPLVGLVNTVLKHTVILKVTANILQLOVPS
ANDELIVKIPLDWAGNTPLVKTIYEFHMTTEOATIRMDTSASGPTRLSDCAT
SHGSLRIQLHLKLSFLVNAKOVNLLVPSLPLNLYKNQLCPVIEASNGMYADLLQ
VKPISLISIDLEFDLXPALKGDTIOLYGLAKLDSOGKVTNKNMNSASLTMPDLD
NIPSLIYSDVVKAAVAAYLSPEEFMLDSVLPRESHRLKSSIGLNEKAAXKLS
TOYIKIILQDTPPEFPIIDGHAQKQVLYLEVPSEEARLPLELTGIESSEAOFTKG
DQILINLINSISDRLOMNSGIMFQDPVPLNITTEIILHILPNKRLSGVPSVL
VKALGFENAESLSLSDALVLTLPASLMKPTSPVSO"

BASE COUNT

419 a 493 c 414 g 358 t

ORIGIN

Query Match 80.7%; Score 1355; DB 9; Length 1684;
Best Local Similarity 99.8%; Pred. No. 0;

Matches 1625; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 53 ggaagggccgagagctcagcgctcccaagctcgcacccgactcgtccctctga 112
DB 6 GGAAGGGCCGAGAGCTCCAGCGTGCAGGCTGCGATCCTGCACTTGCCTCCCTGA 65
QY 113 caacttgggaagaatggccgcccgttgacattcaacctctctgtgttctgagcga 172
DB 66 CACCTGGGAAGATGGCCGGCCCGTGGACCTTCACCCCTTCTGCTGTTGGCGGACCA 125
QY 173 ccttgatccaagcaccctcagtcaccaatgcaatctcaatcctcggcccaaatcaca 232
DB 126 CCTTGATCCAAACCCCTCACTCCCACTGACGTTCTCATCTCGGCCCAAAAGTCATCA 185
QY 233 aagaaagctgacacagagctgaaagacacacagccacagagatcctcgacagcgtgc 292
DB 186 AAGAAAGCTGACACAGGAGCTGAAGGACACACAGCCACGACATCCGACAGACTGC 245
QY 293 cgcgtcagtgccatgcyggaagaaagccagccgagag-attcctgtgctggcagcctg 351
DB 246 CGCTGCTAGTGCATGGCGGAAAGCCAGCGGAGGCAATCCCTGTGCTGGGCGAGCTGG 305
QY 352 tgaacacgctcctgaagcacaatcctgctgaaagtgcaatcaagctaaacatcctcagc 411
DB 306 TGAACACCGTCTCTGAAGCAGCTCACTGTGCTGAAGGTCAACAGCTAATCATCTCCACG 365
QY 412 tgcaggtgaagccctcgagcaatgacagagagctgctgtaagatccccctggagatgg 471
DB 366 TGCAGGTGAAGCCCTGCGCAATGACACAGACTGCTAATCAAGATCCCTCGGACATGG 425
QY 472 tggctgtagtcaacacgcccctgtgtcaagaacatcgttgagtgtccaatgagacgtgag 531
DB 426 TGGCTGAGATTCAACAGCCCTGTGTCAACACCATCTGTGAGTTCCACATGAGACGTGAGG 485
QY 532 cccaagccaacatcgcgatggaacacagtgcaagtggcccaaccgctgtgtcctcaatg 591
DB 486 CCCAAGCCACCATCCGATGAGACACAGTGAAGTGGCCCAACCGGCTGTGCTCTCAGTG 545

QY 592 actgtgcccaagccatgggaagcctgagcatccaactgtgataaagctctctctcgtg 651
DB 546 ACTGTGCCACCGACCGCTGGAGCGCTGCGCATCAATGCTGCAATGACTCTCTCTGG 605
QY 652 tgaagccttaagcaagcagatgaaacctcctagtgcatccctgcccacatctagta 711
DB 606 TGAAGCCTTACTAAGCAAGGTGATCAACCTCTGATGTCATCCCGCCCATCTAGTCA 665
QY 712 aaaaacagctgtgtcccgatgacagagctccttcaatggcatgtatgagaacctctgc 771
DB 666 AAAACACGCTGTCTCCCGTGATGAGGCTTCCCTCAATGATGATGACAGACCTCTGC 725
QY 772 agctgtgagagtgcccatctccatgaagatgagacgctgtgagtttgacctctgtac 831
DB 726 AGCTGTGAAGTGTCCCATTTCCCTAGAGATGACCGTGTGAGTTGACCTTGTATTC 785
QY 832 ctgcacatcaagggtgacacattcagctcattcctcgtgggccaagtgtgtgactcaca 891
DB 786 CTGCATCAAGGGTGAACACATTCACCTTACTGGGGGCCAAGTTGTTGACATCGACG 845
QY 892 gaaaggtgaccaaagtgttcaatlaactctgcagcttcctgacaatgcccacacctgaca 951
DB 846 GAAAGGTGACCAAGTGTCAATTAACCTTGACAGCTTCCCTGACAAATGCCACCTGACA 905
QY 952 acatcccgctcagccctcattcgtgagtcagagacgtggtagaagctgagtgctgtgtgc 1011
DB 906 ACATCCCGTTCAGCCCTCATGCTGAGTCAGACAGCGTGTGAAGCTGAGTGGCTGTGC 965
QY 1012 tctctcagaagaatcaatgactgtgtgtgactctgtgtcctctgagagtgcccatcgcg 1071
DB 966 TCTCTCAGAAAGATTCATGATGCTCTTGTGACTGTGTGCTCTCTGAGATGGCCATCGG 1025
QY 1072 tgaagtcaagcatcgggtgtgataatgaaaggctgacagataagctggatctaccaga 1131
DB 1026 TGAAGTCAAGCATCGGGCTGATCAATGAAGCTGCAATGAAGCGGATGATCACCA 1085
QY 1132 tctgtgaagatccttaactgaaggacactccgagttttttaacaaaggccatggcaag 1191
DB 1086 TCTGTGAAGTCTCTAATCTGACGACACTCCGAGTTTTTTTATGACCAAGGCATGCAAG 1145
QY 1192 tggcccaactgactcgtgtcgtgaagtgtttccctcagtgaaagccctcgcacctgtca 1251
DB 1146 TGGCCCAACTGATCGTGTGGAAGTGTTCCTCCAGTGAAGCCCTCCGCCCTTGTTC 1205
QY 1252 ccttggtgcatggaagcagctcggaaagctcaagttttaaccacaaaggtgacaaactatc 1311
DB 1206 CCTGGGCAATGACAGGACAGCTCGGAAGCTCAGTTTATACACAAAGGTGACCAACTATAC 1265
QY 1312 tcaacttgataatcagactcgtgactcagatccagatcagatgtaactctggatgtgtgt 1371
DB 1266 TCAACTTGAAATACATCACTCTGATCGATCCAGCTTGATGAATCTTGGGATTTGGCTGG 1325
QY 1372 tccaactctgatttctgtaaaaaacatcatcaactgagataatcaactcactcgtgtcga 1431
DB 1326 TCCAACCTGATGTTCGAAAAAACATCATCATGAGATCATTCACCTCATCTGCTGCCGA 1385
QY 1432 accaagaatgcaaatgaatctgggttcccaagtgtcatctgtgtgaaagccttggagattcg 1491
DB 1386 ACCAGATGAGCAAAATTAAATCTGGGGTCCCAAGTGTCAATTTGGGAAGGCTTGGGATTTGG 1445
QY 1492 aggaagctgagatcctcagcagcaagaagtgtccctgtgttactccaagcctctgtgga 1551
DB 1446 AGGAGCTGAGTCTCTACTGAGCAAGAGATGCCCTTGTGCTTACTCCAGCTCTCTTGTGA 1505
QY 1552 aaaccagctctcgtctcctcagtgaaagcttgatgtagcagcatcaaggaaagcctgggtc 1611
DB 1506 AACCCACCTCTCTGCTCTCCAGTGAAGACTTGGATGAGGACCATAGGGAAGGCTGGGT 1565
QY 1612 cccagctggagtatggtgtgagctctatagaacatccctctctgcaatcaataaacac 1671
DB 1566 CCCAGCTGGAGTATGGGTGAGCTCTATAGACCATCCCTCTCTGCAATCAATAAACAC 1625

```

QY 1672 ttgctctgtg 1680
      |||||
Db 1626 ttgctctgtg 1634

RESULT 5
LOCUS AX301908 1035 bp DNA Linear PAT 30-NOV-2001
DEFINITION Sequence 15 from Patent WO0174851.
ACCESSION AX301908
VERSION AX301908.1 GI:17382958
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Majumder K., Spaderna S.K., Taupier R.J., Padigar M.,
Burgess C.E., Shinkets R.A., Spytek K.A., Liu X., Patturajan M. and
Gusev V.Y.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0174851-A 15 11-OCT-2001;
Curagen Corporation (US)
FEATURES
source Location/Qualifiers
1..1035
CDS /organism="Homo sapiens"
/db_xref="taxon:9606"
79..1035
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD13090.1"
/db_xref="GI:17382958"
/translation="MAGPWFTLLCGLLAATLIQATLSPTAVLILGPKYIKETLQEL
KDNHATSILOQLPLDSAMREKPAIGIPVLSIVNTYIKHIMLVKVTANILQLOVPS
ANDDELIVKIPIDMVGAFNPTIVKTIIVEFHTTAEQATITMDTSAGSPRIYSDCAT
SHGSLRIOLHLKLSFLVNAIAKOVNMLVPSLIVKNOCLPPTASFPNGVADLLQ
VKGRCALSPTFSFTTELASRPGKVTKFNNSAASLTMPIDNTIPFSLIVSODVYKAA
VAAVLSPEEFVLLDSVNLSTRORIGPFRPHRNFLNTGCP"
BASE COUNT 230 a 333 c 261 g 211 t
ORIGIN

Query Match 33.5%; Score 562; DB 6; Length 1035;
Best Local Similarity 99.9%; Pred. No. 1e-303;
Matches 682; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 102 ctgccctctgacacctggygaagatgagccgctggaaccttcaacctctctgtgttt 161
      |||||
Db 57 ctgccctctgacacctggygaagatgagccgctggaaccttcaacctctctgtgttt 116

QY 162 gctggagcaaccttgatcaagaacacctcagctccactgcagcttcatctctggccc 221
      |||||
Db 117 gctggagcaaccttgatcaagaacacctcagctccactgcagcttcatctctggccc 176

QY 222 aaaaatcatcaaaagaactgacacagagagctgaagaacacacagcaccagcatct 281
      |||||
Db 177 aaaaatcatcaaaagaactgacacagagagctgaagaacacacagcaccagcatct 236

QY 282 gcaagagctggcgcctgctgaagtgccatgaggggaaagccagccgagag-attccctgtact 340
      |||||
Db 237 gcaagagctggcgcctgctgaagtgccatgaggggaaagccagccgagagattccctgtact 296

QY 341 gggcagcctgtgtgaacacgctctgaagacacatcatctgtcgtgaagatcatcaagctaa 400
      |||||
Db 297 gggcagcctgtgtgaacacgctctgaagacacatcatctgtcgtgaagatcatcaagctaa 356

QY 401 catctccagcctggagaggtgaagccctcgccaaatgaacagagagctgtctaatgaatccc 460
      |||||
Db 357 catctccagcctggagaggtgaagccctcgccaaatgaacagagagctgtctaatgaatccc 416

QY 461 cctgtgaacatgtgtgtgtatltcaaacagccctgtgtcaagacacatcgtgtgttccatc 520
      |||||
Db 417 cctgtgaacatgtgtgtgtatltcaaacagccctgtgtcaagacacatcgtgtgttccatc 476

```

```

QY 521 gacagctgaaggcccaagccacacatccgcatgagacacagctgaagtgagcccaagccgct 580
      |||||
Db 477 gacagctgaaggcccaagccacacatccgcatgagacacagctgaagtgagcccaagccgct 536

QY 581 ggtccatgagctgtgtccacacagcatgggagagctgtgcacccaactgtctgataagct 640
      |||||
Db 537 ggtccatgagctgtgtccacacagcatgggagagctgtgcacccaactgtctgataagct 596

QY 641 ctctctcctgtgtgaagcccttaagcagagatgaagcctctctgtgtccatccctgccc 700
      |||||
Db 597 ctctctcctgtgtgaagcccttaagcagagatgaagcctctctgtgtccatccctgccc 656

QY 701 caatctagtgaaaaaacacagctgtgtccctgtatcgagagctctctcaatgagcatgtgc 760
      |||||
Db 657 caatctagtgaaaaaacacagctgtgtccctgtatcgagagctctctcaatgagcatgtgc 716

QY 761 agaacctctgcaagctgtgtgaag 783
      |||||
Db 717 agaacctctgcaagctgtgtgaag 739

RESULT 6
LOCUS AX301906 963 bp DNA Linear PAT 30-NOV-2001
DEFINITION Sequence 13 from Patent WO0174851.
ACCESSION AX301906
VERSION AX301906.1 GI:17382956
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Majumder K., Spaderna S.K., Taupier R.J., Padigar M.,
Burgess C.E., Shinkets R.A., Spytek K.A., Liu X., Patturajan M. and
Gusev V.Y.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0174851-A 13 11-OCT-2001;
Curagen Corporation (US)
FEATURES
source Location/Qualifiers
1..963
CDS /organism="Homo sapiens"
/db_xref="taxon:9606"
1..963
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD13089.1"
/db_xref="GI:17382957"
/translation="MAGPWFTLLCGLLAATLIQATLSPTAVLILGPKYIKETLQEL
KDNHATSILOQLPLDSAMREKPAIGIPVLSIVNTYIKHIMLVKVTANILQLOVPS
ANDDELIVKIPIDMVGAFNPTIVKTIIVEFHTTAEQATITMDTSAGSPRIYSDCAT
SHGSLRIOLHLKLSFLVNAIAKOVNMLVPSLIVKNOCLPPTASFPNGVADLLQ
VKGRCALSPTFSFTTELASRPGKVTKFNNSAASLTMPIDNTIPFSLIVSODVYK
AAVAAVLSPEEFVLLDSVNLSTRORIGPFRPHRNFLNTGCP"
BASE COUNT 217 a 314 c 237 g 195 t
ORIGIN

Query Match 24.1%; Score 405; DB 6; Length 963;
Best Local Similarity 100.0%; Pred. No. 1.7e-215;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 ggtgaaggtatcatcagcttaacatccctcagcctgagctgaagccctggccaatgacc 438
      |||||
Db 263 ggtgaaggtatcatcagcttaacatccctcagcctgagctgaagccctggccaatgacc 322

QY 439 agagagctgtcagtgaaagatccctctggaactgtgtgtgtatcaaacagccctgtgtca 498
      |||||
Db 323 agagagctgtcagtgaaagatccctctggaactgtgtgtgtatcaaacagccctgtgtca 382

QY 499 agaacatcgtgtgagttcacatgaagcactgaagcccaagccacatccgcatgagacca 558
      |||||

```

Db 383 AGACCATCGTGGAGTTCCATGACGACACGAGCCCAACCATCCGATGACACCA 442
 Oy 559 gtgcaagtgtgcccaccgccctgtgctcctgaagtgtgtccaccagccatggagccttc 618
 Db 443 GTGCAAGTGGCCCCACCCCGCTGTCTCAGTACTGTCCACACGACCAATGGAGGCTTC 502
 Oy 619 gcatccaactgtgtcataagctctctcctcctgtgtgaacgcttaagcagatcatga 678
 Db 503 GCATCCAACTGCTGCATTAAGCTCTCTCTCCTGTGTAAGCCCTTACGTAAGAGTCATCA 562
 Oy 679 acctcctagtgtccatccctgtcccaatctagtgaaaaacacagctgtgtccgtgacag 738
 Db 563 ACCCTCTAGTGGCAATCCCTGGCCCAATCTAGTGAATAAACACAGCTGTCTCCCGTATCGAAG 622
 Oy 739 cctcctcaatgcatgtatgatgacagcctcctcagctggtgaag 783
 Db 623 CTTCTCTTCAATGGCATGTATGTCAGACCTCTCTGACAGCTGTGAAG 667

RESULT 7
 AL355392
 LOCUS
 DEFINITION Human DNA sequence from clone RP5-1187J4 on chromosome 20q11.1-11.23 Contains ESTs, STSs, GSSs and two CpG islands. Contains the gene for novel protein similar to mouse von Ebner salivary gland protein, the gene for a novel protein similar to rat RYF3, the LOC51654 gene for a novel protein (GGI-05) similar to rat CDK5 activator-binding protein and the SNTA1 gene for alpha syntrophin (dystrophin-associated protein A1, 59kD, acidic component), complete sequence.
 AL355392 AL117331
 VERSION AL355392.7 GI:10178502
 KEYWORDS HTG; CGI-05; CpG island; RYF3; SNT1; SNTAL; syntrophin.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 159440)
 AUTHORS Bird,C.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 COMMENT On May 14, 2001 this sequence version replaced gi:6982468 gi:9213549.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

This sequence is the entire insert of clone RP5-1187J4 The true left end of clone RP11-327D19 is at 62286 in this sequence. The true right end of clone RP11-49G10 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats: all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-1187J4 is from the library RPCR-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

FEATURES

VECTOR: pCPYPAC2.
 Location/Qualifiers
 1..159440
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="q11.1-11.23"
 /clone="RP5-1187J4"
 /clone_1ib="RPI-5"
 1535..1848
 /note="157 copies 2 mer ga 61% conserved"
 1757..1824
 /note="Random repeat. weak data"
 join(3408..3471,5782..5867,7116..7295,8242..8295,8799..8957,9841..9908,10729..10774,11885..11948,12794..12870,15591..15761)
 /gene="dJ1187J4.1"
 /note="match: ESTs: Em:BE933124 Em:BE061388"
 /product="dJ1187J4.1.1 (novel protein similar to mouse von Ebner salivary gland protein, isoform 1.)"
 /evidence=not_experimental
 3408..15761
 /gene="dJ1187J4.1"
 join(3408..3471,5782..5867,7116..7295,8242..8295,8799..8957,9841..9908,10729..10774,11885..11948,12794..12870,15591..15650)
 /gene="dJ1187J4.1"
 /note="Continues in Em:AL121901 as bA49G10.6"
 /codon_start=1
 /evidence=not_experimental
 /product="dJ1187J4.1.1 (novel protein similar to mouse von Ebner salivary gland protein, isoform 1.)"
 /protein_id="CAC34050.1"
 /db_xref="GI:13274680"
 /translation="LCPIVEASFNGMTVDLQYKVPISLSIDREFDLIPAIKGDITQLYGAALDSQKRVTFMNFNSASLTPITDNIPIFSLIVSQDVVKAVAVLSPEEFMVLIDSVLPESAHRLKSGISGLINERKADKLSGTQIVKILQDPTEPFIIDGSHAKVAQLIVLEVPSSSEALRPDLFTLTGTEASEAQPTFGDQLIINLNISSDRILNMISGIMFQPDVILKNITETIHSLILPNQNGKRSRSPVSLVALDFEAESLITDALVLPASLWKPSSPVSO"
 3817..3915
 /gene="dJ1187J4.1"
 /note="match: STS: Em:G10242"
 complement(join(3917..4006,31119..31326))
 /note="match: STS: Em:U75536"
 3917..4143
 /gene="dJ1187J4.1"
 /note="match: STS: Em:G10242"
 4662..4691
 /note="15 copies 2 mer tt 86% conserved"
 join(5758..5867,7116..7295,8242..8295,8799..8957,8841..9908,10729..10774,11885..11948,12794..12870,15591..15761)
 /gene="dJ1187J4.1"
 /note="match: CDNAS: Em:U46068
 match: ESTs: Em:A1924632 Em:AW63590 Em:AW384404 Em:AW449208 Em:AW167650 Em:AW381927 Em:AW167610 Em:AW577643 Em:AW176323 Em:A1024451 Em:DA5553 Em:AW176329 Em:AW176327 Em:AW176325 Em:A1909965 Em:A1909964 Em:AA428580 Em:AW384405"
 /product="dJ1187J4.1.2 (novel protein similar to mouse von Ebner salivary gland protein, isoform 2.)"
 /evidence=not_experimental
 complement(6761..7253)
 /note="match: GSS: Em:A0149447"
 join(7182..7295,8242..8295,8799..8957,9841..9908,10729..10774,11885..11948,12794..12870,15591..15650)
 /gene="dJ1187J4.1"
 /codon_start=1
 /evidence=not_experimental
 /product="dJ1187J4.1.2 (novel protein similar to mouse von Ebner salivary gland protein, isoform 2.)"
 /protein_id="CAC15880.1"

misc_feature
 misc_feature
 misc_feature
 repeat_region
 mRNA
 misc_feature
 CDS
 misc_feature
 CDS


```

/db_xref="GI:11137741"
/translation="MPTIDNPISLVSDVYKAAVAVLSPFPMWLLDSVLPESAH
RKSKSLGLEKAKDKIGSTQIYKILITQDIPFPIDQGAQVAVQLIVLEVPSSEAR
PLPLFLGTFASAEQFTYKGDQLITLNNTNSDRQLMNSGIGWQPPDLKNIITFIIH
SILPLPNQNGKLRSGVPVSLTKALGFVAEESLTKDALVLTASLTKPSSPVSQ"
10970..11009
/note="20 copies 2 mer ac 85% conserved"
misc_feature
complement(11466..11766)
/note="match: GSS: Em: A0606066"
complement(11631..11769)
/note="match: GSS: Em: A0427662"
11766..12192
/gene="dJ1187J4.1"
/note="match: GSS: Em: A0403818"
11768..12108
/gene="dJ1187J4.1"
/note="match: GSS: Em: A0373301"
11775..12108
/gene="dJ1187J4.1"
/note="match: GSS: Em: A0662246"
14322..14833
/note="256 copies 2 mer ca 68% conserved"
14836..14871
/note="18 copies 2 mer ac 88% conserved"
complement(15031..15459)
/note="match: GSS: Em: A0694930"
15736..15741
/gene="dJ1187J4.1"
15754
/gene="dJ1187J4.1"
15761
/gene="dJ1187J4.1"
complement(17559..18000)
/note="match: GSS: Em: A0349609"
complement(17617..18004)
/note="match: GSS: Em: A0350711"
18109..18565
/note="match: STS: Em: HS81207T"
complement(20661..21112)
/note="match: GSS: Em: A0019396"
complement(20698..21115)
/note="match: GSS: Em: A0178024"
21105..21680
/note="match: STS: Em: HS156422T"
complement(22577..22968)
/note="match: GSS: Em: A0508829"
22977..23573
/note="match: GSS: Em: A0281488"
complement(37780..38128)
/note="match: STS: Em: HSPF13P11"
complement(39112..39718)
/note="match: GSS: Em: A0636647"
complement(39266..39746)
/note="match: GSS: Em: A0827010"
39747..40042
/note="match: GSS: Em: A0282137"
39754..40132
/note="match: GSS: Em: A0252720"
39763..40037
/note="match: GSS: Em: A0029439
match: STS: Em: GS0113"
complement(39950..40165)
/note="match: GSS: Em: AL207848"
complement(39950..40334)
/note="match: GSS: Em: A0556676"
39961..40217
/note="match: GSS: Em: A0143633"
complement(39964..40169)
/note="match: STS: Em: G02503"
complement(39991..40409)
/note="match: GSS: Em: A0596761"
complement(40002..40172)
/note="match: GSS: Em: AL185783"

```

```

misc-feature      /note="match: GSS: Em: A0052600"
Query Match      10.8%; Score 181; DB 9; Length 159440;
Best Local Similarity 100.0%; Pred. No. 9.9e-90;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 869 ggcgaagttcttgactcacaggaagtgaccagtggttcataaactctgcagcttc 928
DB 7115 ggccagagttgttgactcacaggaagtgaccagtggttcataaactctgcagcttc 7174
QY 929 cctgcacaatgcccacccctggaacaacatcccggttcagcctccatcgtgagtcaggaagctggt 988
DB 7175 ccgacgaatgcccacccctggaacaacatcccggttcagcctccatcgtgagtcaggaagctggt 7234
QY 989 gaagctgcagtcagtcgcctcgtctctctccagaagaattcatggttcctgttgactctgt 1048
DB 7235 gaagctgcagtcagtcgcctcgtctctctccagaagaattcatggttcctgttgactctgt 7294
QY 1049 g
DB 7295 g 7295

RESULT 8
LOCUS      HSBAA49G10
DEFINITION Human DNA sequence from clone 161593 bp. DNA linear PRI 20-JUN-2001
HBBAA49G10
LOCUS      HSBAA49G10
DEFINITION Human DNA sequence from clone 161593 bp. DNA linear PRI 20-JUN-2001
a gene similar to bovine salivary protein BSP30, the LOC51297 gene
for LUNX protein, 5' end of a gene encoding a protein similar to
murine von esner minor salivary gland protein, a novel gene, a
pseudogene similar to STAT-induced STAT inhibitor-2, a pseudogene
similar to ribosomal protein l12, a putative novel transcript,
ESTs, STSs, GSSs and a cpg island, complete sequence.
AL121901
AL121901.20 GI:8249854
HTG: BSP30; CPG island; LOC51297; LUNX; RPL12; STAT-inhibitor.
human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 161593)
Tracey/A.
Direct Submission
Submitted (14-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
On Jun 5, 2000 this sequence version replaced gi:8017404.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; SW:, SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
IMPORTANT: This sequence is not the entire insert of clone
RP11-49G10 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-49G10 is at 1 in this sequence.
The true left end of clone RP5-1187J4 is at 161494 in this sequence.
The true right end of clone RP4-7330J3 is at 27823 in this
sequence. This sequence was finished as follows unless otherwise
noted: all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., phred
quality >= 30); an attempt was made to resolve all sequencing

```


problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. Rpl1-49G10 is from the library RPCR-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.
Location/Qualifiers

```

FEATURES
    source
        1..161593
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="20"
        /clone="Rpl1-49G10"
        /clone_1lb="RPCI-11.1"
        complement(1..107)
        /note="match: GSS: Em: A0607627"
        6..706
        /note="match: GSS: Em: A0051854"
        137..548
        /note="match: GSS: Em: A0550113"
        530..603
        /note="37 copies 2 mer ca 66% conserved"
        533..686
        /note="2 copies 77 mer 83% conserved"
        613..678
        /note="33 copies 2 mer ca 75% conserved"
        3030..3174
        /note="L1P93 repeat: matches 5941..6084 of consensus"
        5140..5183
        /note="22 copies 2 mer tt 75% conserved"
        5184..5344
        /note="L1PA2 repeat: matches 5986..6146 of consensus"
        complement(6104..6463)
        /note="match: GSS: Em: A0339300"
        6120..6239
        /note="2 copies 60 mer 85% conserved"
        complement(8476..9002)
        /note="match: GSS: Em: A0537190"
        complement(14766..15061)
        /note="match: STS: Em: H8922114S"
        14772..14779
        /note="Random repeat. Forced join. Approximately 500 bases missing according to restriction digest."
        16219..16387
        /note="match: GSS: Em: AL001671"
        16225..16387
        /note="match: GSS: Em: AL013146"
        18325..18826
        /note="match: GSS: Em: A0458133"
        18328..18828
        /note="match: GSS: Em: A0532957"
        18333..18770
        /note="match: GSS: Em: A0765930"
        complement(20975..21245)
        /note="match: GSS: Em: A2102442"
        20999..21238
        /note="4 copies 60 mer 69% conserved"
        21031..21263
        /note="match: GSS: Em: A2055779 Em: A2084220"
        21144..21225
        /note="41 copies 2 mer ag 75% conserved"
        complement(21172..21428)
        /note="match: GSS: Em: A0395327"
        complement(23473..23950)
        /note="match: GSS: Em: A0727877"
        complement(23637..23948)
        /note="match: GSS: Em: A0691469"
        23950..24355
        /note="match: GSS: Em: A0152593"
        23955..24456
        /note="match: GSS: Em: A0508832"
        26979..27422
        /note="match: GSS: Em: A0636646"

```

```

repeat_region
    31882..32061
    /note="3 copies 60 mer 73% conserved"
repeat_region
    31921..32032
    /note="56 copies 2 mer ct 63% conserved"
repeat_region
    33693..33762
    /note="MER47 repeat: matches 2251..2322 of consensus"
    33733..33785
    /note="MER47 repeat: matches 1..55 of consensus"
    join(35535..35585,36507..36678,40308..40452,41455..41562,42783..42935,45604,46980..47043,47896..47936)
    /note="MER47 repeat: matches 1..55 of consensus"
    48600..48787
    /gene="DA49G10.1"
    /note="match: CDNAS: Em: U79414 Em: U79413 Em: AX061621 match: ESTs: Em: A1654622"
    /product="DA49G10.1 (similar to bovine salivary protein BSP30)"
    /evidence="not_experimental"
    35335..48787
    /gene="DA49G10.1"
    join(36522..36678,40308..40452,41455..41562,42783,45523..45604,46980..47043,47896..47936)
    /gene="DA49G10.1"
    /note="match: proteins: Tr: P79125 Sw: P07743"
    /codon_start=1
    /evidence="not_experimental"
    /product="DA49G10.1 (similar to bovine salivary protein BSP30)"
    /protein_id="CAC03546.1"
    /db_xref="GI:9801234"
    /transcript="M10LMKLVLLCGVLTGTSSESLDNLGNDLSNVVDKLEPVLHEGL ETVDDTKLEKLEKLDVGLGKSSAMOLAKOAEAKLNNVSKLLPNTDIFGL KINSLLIDVKAEPIDOGKIGNSPPVANTVAGPIGIIINLKASLDLTAATIER DPQHQPVAVGECASDPTSLSLDHSQIINKFNVSINTLKSTVSSLLQKEIC LRIETHSLDVAVNIDVVDNPDQHKQDQTLT"
    38155..38751
    /note="L1PA8 repeat: matches 4514..5125 of consensus"
    38750..39785
    /note="L1PA8 repeat: matches 5128..6163 of consensus"
    48769..48774
    /gene="DA49G10.1"
    48787
    /gene="DA49G10.1"
    52540..53450
    /note="CpG island"
    /evidence="not_experimental"
    complement(53851..54131)
    /note="match: GSS: Em: A0897745"
    53831..54141
    /note="match: STS: Em: G31731"
    53842..53969
    /note="match: GSS: Em: A0429347"
    complement(53844..54141)
    /note="match: GSS: Em: AF101963"
    53858..54022
    /note="match: GSS: Em: A0392891"
    complement(53908..54141)
    /note="match: GSS: Em: A0415287"
    53958..54139
    /note="match: GSS: Em: A0378081"
    53970..54139
    /note="match: GSS: Em: A0527291"
    53972..54142
    /note="match: GSS: Em: A0429347"
    54000..54141
    /note="match: GSS: Em: A0392510"
    complement(54027..54139)
    /note="match: GSS: Em: A0533658"
    54030..54138
    /note="match: GSS: Em: A0600935"
    54041..54149
    /note="match: GSS: Em: A0392891"
    54767..55668
    /gene="DA49G10.2"

```

REFERENCE
AUTHORS

1 (bases 1 to 89536)
Mizuy,D.M., Adams,C., Adio-Oduola,B., Ali-Usman,F.R., Allen,C.,
Aisbrooks,S.L., Amaralunga,H.C., Aye,J.R., Banks,T., Barbarta,J.,
Benton,T., Bimige,K., Blakenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brilewa,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudat,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Loulsgaard,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metker,M., Miner,G., Miner,Z., Mitchell,T., Mohapatra,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nockenko,S.,
Ogulu,M., Okwomu,G., Otagunye,N., Oyiedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojudoakan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N.,
Slisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swatek,A., Tabors,P., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telirod,B., Thomas,N.,
Thomas,S., Usmali,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Wallington,S., Williams,G., Williamson,A., Wlaczky,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Unpublished
2 (bases 1 to 89536)
Worley,K.C.

Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15636537.

----- genome center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GAEB
Center clone name: CH230-2K16

----- Summary Statistics

Assembly Program: Phrap; version 0.990329first call to
findPhraplist

Consensus quality: 73144 bases at least Q40
Consensus quality: 80017 bases at least Q30
Consensus quality: 85368 bases at least Q20
Estimated insert size: 69357; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafc_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
4714: contig of 4713 bp in length
4813: gap of unknown length
4814: contig of 2629 bp in length
7443: gap of unknown length
7543: contig of 1428 bp in length
8971: gap of unknown length
9071: contig of 2548 bp in length
11618: contig of 2346 bp in length
11719: gap of unknown length
14064: contig of 2346 bp in length
14164: gap of unknown length
14165: contig of 2434 bp in length
16599: gap of unknown length
16599: contig of 1860 bp in length
18559: gap of unknown length
21636: contig of 2978 bp in length
21736: gap of unknown length
23091: contig of 1355 bp in length
23191: gap of unknown length
24363: contig of 1172 bp in length
24463: gap of unknown length
26616: contig of 2153 bp in length
26716: gap of unknown length
29328: contig of 2612 bp in length
29428: gap of unknown length
30861: contig of 1433 bp in length
30961: gap of unknown length
32666: contig of 1705 bp in length
32766: gap of unknown length
34751: contig of 1985 bp in length
34851: gap of unknown length
36722: contig of 1871 bp in length
36822: gap of unknown length
37990: contig of 1168 bp in length
38090: gap of unknown length
38951: contig of 1761 bp in length
39852: gap of unknown length
41482: contig of 1531 bp in length
41582: gap of unknown length
42970: contig of 1388 bp in length
43070: gap of unknown length
44717: contig of 1647 bp in length
44817: gap of unknown length
46731: contig of 1914 bp in length
46831: gap of unknown length
48721: contig of 1890 bp in length
48821: gap of unknown length
50484: contig of 1663 bp in length
50584: gap of unknown length
52333: contig of 1749 bp in length
52433: gap of unknown length
53718: contig of 1285 bp in length
53818: gap of unknown length
55081: contig of 1263 bp in length
55181: gap of unknown length
56437: contig of 1256 bp in length
56537: gap of unknown length
58395: contig of 1858 bp in length
58495: gap of unknown length
60176: contig of 1681 bp in length
60276: gap of unknown length
62186: contig of 1910 bp in length
62286: gap of unknown length
64498: contig of 2212 bp in length
64598: gap of unknown length
65989: contig of 1391 bp in length
66089: gap of unknown length
67493: contig of 1404 bp in length
67593: gap of unknown length
69641: contig of 2048 bp in length

```

```

* 69642 69741: gap of unknown length
* 69742 70862: contig of 1121 bp in length
* 70863 70962: gap of unknown length
* 70963 72389: contig of 1427 bp in length
* 72390 72489: gap of unknown length
* 72490 74082: contig of 1593 bp in length
* 74083 74182: gap of unknown length
* 74183 75218: contig of 1036 bp in length
* 75219 75318: gap of unknown length
* 75319 76810: contig of 1491 bp in length
* 76810 76909: gap of unknown length
* 76910 78708: contig of 1799 bp in length
* 78709 78808: gap of unknown length
* 78809 79849: contig of 1041 bp in length
* 79850 79949: gap of unknown length
* 79950 81538: contig of 1589 bp in length
* 81539 82748: gap of unknown length
* 82749 82849: contig of 1110 bp in length
* 82849 84058: gap of unknown length
* 84059 84158: gap of unknown length
* 84159 85695: contig of 1537 bp in length
* 85696 85795: gap of unknown length
* 85796 86875: contig of 1080 bp in length
* 86876 86975: gap of unknown length
* 86976 88063: contig of 1088 bp in length
* 88064 88163: gap of unknown length
* 88164 89536: contig of 1373 bp in length.

FEATURES
    source
        1..89536
        /organism="Rattus norvegicus"
        /db_xref="taxon:10116"
        /clone="CH230-2K16"

BASE COUNT      21918 a 20067 c 20357 g 22306 t 4888 others

ORIGIN
Query Match      1.5%; Score 26; DB 2; Length 89536;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      456 atccccctgacatgctgctgalt 481
Db      23604 ATCCCCCTGCAGACATGCTGCCTGATT 23579

RESULT  11
MMU46068      1629 bp  mRNA  linear  ROD 10-AUG-2000
LOCUS
DEFINITION Mus musculus von Ebner minor salivary gland protein mRNA, complete
cds.
ACCESSION U46068
VERSION U46068.2 GI:9789706
KEYWORDS house mouse strain-Swiss-Webster.
SOURCE
ORGANISM Mus musculus
    Mus musculus
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
        1 (bases 1 to 1629)
        Sneed,M.L., Villanueva,J., Paine,M.L., Lei,Y.P., Zhu,D.H.,
        Luis,J., Xia,Y.-R. and Yang,J.-N.
        Direct Submission
        Submitted (17-JAN-1996) Malcolm L. Sneed, Center for Cranio
        Molecular Biology, University of Southern California, 2250 Alcazar
        St., Los Angeles, CA 90033, USA
        On Aug 10, 2000 this sequence version replaced gi:1184789.
        Location/Qualifiers
            1..1629
            /organism="Mus musculus"
            /strain="Swiss-Webster"
            /db_xref="taxon:10090"
            /chromosome="2"
            /map="2; D2Mit19n and D2Mit25n"
FEATURES
    source

```

```

CDS
58..1482
/codon_start=1
/product="von Ebner minor salivary gland protein"
/protein_id="AAA87581.2"
/db_xref="GI:9789707"
/translation="MAGPITITLGLGATLVQANVYPVAVNLGPEVIOKHTQAL
KHQDPAITLQELPLRAMODKSGIPILDSSEVAVLKIIIMKVTSAITLQDVPST
VNOELVARIPLDMVAGNLPLIKTIVEFQSTVEVALIKRVRSSGPAHNLSDCSN
ESTRLSLHLKLSFVNSLAKNVMNLVPLAPOLYKNNLCPMIOADDMYEDRLT
TAPILSPGALFEGLSPAIDSNILNLKKKLLDSARVYNNNSATSIEMETTPG
APFSLVRQDLYNALVTWLVPRKEVILLRPVITDVANQLOMDIKETNAEKANLIGPT
OMLKTFSTPHIVLNEGASARAOSVILEPPTNDVRFESLGEISYKQPTEDN
RLMNFNNVSEIRIKMISDILKLTDEPVNLKTLKILEYLLPNEKGLRTGIPMSMS
KALGKRMMSVSGALKLTPASS"

BASE COUNT      411 a      497 c      382 g      339 t

Query Match
Best Local Similarity 1.4%; Score 24; DB 10; Length 1629;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 atccccctgacatggtgcctga 479
|||||
Db 388 ATCCCCCTGACATGCTGCTGCA 411

RESULT 12
AP000494
LOCUS      AP000494      200000 bp      DNA      linear      PRI 28-SEP-1999
DEFINITION Homo sapiens genomic DNA, chromosome 3p21.3, clone:301 to 308,
            anti-oncogene region, section 2/5.
ACCESSION  AP000494
VERSION     AP000494.1 GI:5926671
KEYWORDS
SOURCE      Homo sapiens DNA.
            Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (sites)
            Nakamura,Y., Isomura,M., Daigo,Y., Tamari,M. and Ishikawa,S.
            DNA sequence analysis of a 1.2-Mb region on chromosome 3p21.3
            Published only in Database (1999) In press
            2 (bases 1 to 200000)
            Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
            Direct Submission
            Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Miki
            Hirakawa, Japan Science and Technology Corporation (JST), Advanced
            Databases Department, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
            Japan (E-mail:mikae@tokyo.jst.go.jp,
            URL:http://www-alls.tokyo.jst.go.jp/, Tel:81-3-5214-8470,
            Fax:81-3-5214-8470)
            This sequence is conducted by Japanese Foundation for Cancer
            Research as a JST sequencing team.
            Principal Investigator: Yusuke Nakamura Ph.D
            Phone:+81-3-5449-5372, Fax:+81-3-5449-5433,
            yusuke@nc.ims.u-tokyo.ac.jp
            The sequence is submitted by Human Genome Sequencing in ALLS
            project of JST
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
            For further information about this sequences, please visit our
            sequence archive Web site (http://www-alls.tokyo.jst.go.jp/HGS/top.
            html) or send email to webmaster@www-alls.tokyo.jst.go.jp.
            Location/Qualifiers
            1..200000
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /map="3p21.3"
            /note="301-308"
            complement(112666..112810)
            /note="sts69202:The location is between each flanking site

FEATURES
Source
STS
complement(112666..112810)
/note="sts69202:The location is between each flanking site

of PCR primers."
/db_xref="GDB:4569933"
BASE COUNT      52190 a      42809 c      46061 g      58756 t      184 others

Query Match
Best Local Similarity 1.3%; Score 22; DB 9; Length 200000;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 842 gggagacacatcagctctac 863
|||||
Db 135252 GGGAGACACCATTCAGCTCTAC 135273

RESULT 13
AX301962/c
LOCUS      AX301962      21 bp      DNA      linear      PAT 30-NOV-2001
DEFINITION Sequence 69 from Patent WO0174851.
ACCESSION  AX301962
VERSION     AX301962.1 GI:17382977
KEYWORDS
SOURCE      synthetic construct.
            synthetic construct.
            artificial sequence.
ORGANISM
REFERENCE   1 (sites)
            Majumder,K., Spaderna,S.K., Taupier,R.J., Padigaru,M.,
            Burgess,C.E., Shmukets,R.A., Spleyk,K.A., Liu,X., Paturajan,M. and
            Gusev,V.Y.
            Novel proteins and nucleic acids encoding same
            Patent: WO 0174851-A 69 11-OCT-2001;
            Curagen Corporation (US)
FEATURES
Source
            1..21
            Location/Qualifiers
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="Ag719 Reverse Primer Sequence"
BASE COUNT      4 a      6 c      5 g      6 t

Query Match
Best Local Similarity 1.2%; Score 21; DB 6; Length 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 aatgacagagcgcgcagtc 452
|||||
Db 21 AATGACGAGAGCTGCTAGTC 1

RESULT 14
AF184920
LOCUS      AF184920      667 bp      mRNA      linear      ROD 05-OCT-1999
DEFINITION Rattus norvegicus GDNF-family receptor alpha 3 (GFRalpha-3) mRNA,
            partial cds.
ACCESSION  AF184920
VERSION     AF184920.1 GI:6010760
KEYWORDS
SOURCE      Norway rat.
            Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 667)
            Stover,T., Gong,T.-W., Cho,Y., Altschuler,R.A. and Lomax,M.I.
            Expression of neuturin, artemin, persephin and their receptors
            GFRa-2 and GFRa-3 in the mature rat cochlea
            Unpublished
            2 (bases 1 to 667)
            Stover,T., Gong,T.-W., Cho,Y., Altschuler,R.A. and Lomax,M.I.
            Direct Submission
            Submitted (13-SEP-1999) Kresge Hearing Research Institute,
            University of Michigan, 9301E MSRB III, 1150 West Medical Center

```

FEATURES Drive, Ann Arbor, MI 48109-0648, USA
Location/Qualifiers

source 1..667
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="cochlea; substantia nigra"
1..>667
/gene="GFRalpha-3"
1..>667
/gene="GFRalpha-3"
/product="GDNF-family receptor alpha 3"
/codon_start=2
/protein_id="AAFO1242.1"
/db_xref="gi:6010761"
/translation="TVTSKPKRMNLSKLSMLKPDSDLCFKPAMCLTLDRCDRLEKAY
GEACGICQRIHCLLAQLRFFEKAEASHAGCLLCPAPEDAGGERRRNTIASCA
LPSVAPNCLDLSFRADPLCRSLMDPOTCHHPMDIGTCATDSRCLRAYLIGT
AMTPFISKNTVTVALGCTCRSGNLQDECEQLEKFSQNPCLMEALIAKMRHRLQ
SODW"

gene
CDS

BASE COUNT 138 a 222 c 175 g 132 t

ORIGIN

Query Match 1.2%; Score 21; DB 10; Length 667;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 tggactccagaccacatgcc 40
|||||
Db 384 tggactccagaccacatgcc 404

RESULT 15
AC096278 32369 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-11L24, *** SEQUENCING IN PROGRESS
ACCESSION AC096278.2 GI:17943975
VERSION AC096278.2
KEYWORDS HTG, HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 32369)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowen,S., Briteva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burth,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garcia,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsi,F., Howard,S., Huber,J., Huix,S., Hume,J.,
Jachson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivel,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguru,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Prins,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojuben,I., Rolle,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,D., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 32369)
Worley,K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15627898.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GEVR
Center clone name: CH230-11L24
----- Summary Statistics
Assembly program: Phrap; version 0.990329Flrst call to
findphraplist
Consensus quality: 25575 bases at least Q40
Consensus quality: 29387 bases at least Q30
Consensus quality: 32762 bases at least Q20
Estimated insert size: 11642; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; average-fp estimation
Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 2641 2640: contig of 2640 bp in length
* 2741 4533: contig of 1793 bp in length
* 4534 4633: gap of unknown length
* 4634 6203: contig of 1570 bp in length
* 6204 6303: gap of unknown length
* 6304 7920: contig of 1617 bp in length
* 7921 8020: gap of unknown length
* 8021 9285: contig of 1265 bp in length
* 9286 9385: gap of unknown length
* 9386 10442: contig of 1057 bp in length
* 10443 10542: gap of unknown length
* 10543 12042: contig of 1500 bp in length
* 12043 12142: gap of unknown length
* 12143 13857: contig of 1715 bp in length
* 13858 13957: gap of unknown length
* 13958 15637: contig of 1680 bp in length
* 15638 15737: gap of unknown length
* 15738 16927: contig of 1190 bp in length
* 16928 17027: gap of unknown length
* 17028 19301: contig of 2274 bp in length
* 19302 19401: gap of unknown length
* 19402 20634: contig of 1233 bp in length

```

* 20635 20734: gap of unknown length
* 20735 21741: contig of 1007 bp in length
* 21742 21841: gap of unknown length
* 21842 23151: contig of 1310 bp in length
* 23152 23251: gap of unknown length
* 23252 24576: contig of 1325 bp in length
* 24577 24676: gap of unknown length
* 24677 25830: contig of 1154 bp in length
* 25831 25930: gap of unknown length
* 25931 27157: contig of 1227 bp in length
* 27158 27257: gap of unknown length
* 27258 28348: contig of 1091 bp in length
* 28349 28448: gap of unknown length
* 28449 29559: contig of 1111 bp in length
* 29560 29659: gap of unknown length
* 29660 30777: contig of 1118 bp in length
* 30778 30877: gap of unknown length
* 30878 32369: contig of 1492 bp in length.

```

FEATURES
Source 1. .32369 Location/Qualifiers

BASE COUNT 8252 a 6711 c 7651 g 7723 t 2032 others
ORIGIN /organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-11L24"

Query Match 1.2%; Score 21; DB 2; Length 32369;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 280 ctgcagcagctgcgcgcctc 300
|||||
Db 22520 CTGCAGCAGCTGCCGCTCCTC 22540

Search completed: September 16, 2002, 06:28:35
Job time: 12193 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 04:24:12 : Search time 390.66 seconds
(without alignments)
7385.444 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680
Sequence: 1 ggtgtgcagatataagtt.....tcaataaacactgtcgtgtg 1680

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq_032802:*
1: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--|
| 1 | 1680 | 100.0 | 1680 | 21 | AA229724 | Human Lung specific Nucleotide sequenc |
| 2 | 1628 | 96.9 | 1680 | 22 | AAH77948 | Human NTAP CDNA cl |
| 3 | 1458 | 86.8 | 1707 | 21 | AA287210 | Human PRO1357 (UNO |
| 4 | 1457 | 86.7 | 1636 | 21 | AAA37057 | Human DNA encoding |
| 5 | 1457 | 86.7 | 1636 | 22 | AA546064 | Human PRO1357 CDNA |
| 6 | 1457 | 86.7 | 1636 | 22 | AAF92096 | Human DNA encoding |
| 7 | 1457 | 86.7 | 1636 | 22 | AA534293 | Human DNA encoding |
| 8 | 1457 | 86.7 | 2036 | 22 | AAH77950 | Nucleotide sequenc |
| 9 | 1064 | 63.3 | 2121 | 23 | AA572707 | DNA encoding novel |

| | | | | | | |
|----|-----|------|------|----|----------|--------------------|
| 10 | 562 | 33.5 | 1035 | 22 | ABA01989 | Human NOV6b coding |
| 11 | 523 | 31.1 | 1177 | 21 | AAF18254 | Lung cancer associ |
| 12 | 522 | 31.1 | 1175 | 21 | AA297023 | Human secreted pro |
| 13 | 436 | 26.0 | 1172 | 21 | AA297114 | Human secreted pro |
| 14 | 420 | 25.0 | 420 | 21 | AAH30554 | Human colon cancer |
| 15 | 405 | 24.1 | 963 | 22 | ABA01986 | Human NOV6a coding |
| 16 | 271 | 16.1 | 465 | 21 | AA274959 | Human ORFX ORF514 |
| 17 | 229 | 13.6 | 232 | 16 | AA221378 | Human gene sRNA |
| 18 | 213 | 12.7 | 423 | 22 | AA118484 | Human breast cance |
| 19 | 213 | 12.7 | 462 | 22 | AA109612 | Human breast cance |
| 20 | 181 | 10.8 | 594 | 22 | ABA61288 | Human foetal liver |
| 21 | 181 | 10.8 | 594 | 22 | ABA29112 | Probe #7578 for ge |
| 22 | 181 | 10.8 | 594 | 22 | AAK09583 | Human brain expres |
| 23 | 181 | 10.8 | 594 | 22 | AAK35477 | Human bone marrow |
| 24 | 181 | 10.8 | 594 | 22 | AA117050 | Probe #6983 for ge |
| 25 | 181 | 10.8 | 594 | 22 | AA141191 | Probe #9877 used t |
| 26 | 180 | 10.7 | 180 | 22 | ABA73788 | Human foetal liver |
| 27 | 180 | 10.7 | 180 | 22 | ABA38963 | Probe #17429 for g |
| 28 | 180 | 10.7 | 180 | 22 | AAK22236 | Human brain expres |
| 29 | 180 | 10.7 | 180 | 22 | AAK48404 | Human bone marrow |
| 30 | 180 | 10.7 | 180 | 22 | AA126228 | Probe #16161 for g |
| 31 | 180 | 10.7 | 180 | 22 | AA154231 | Probe #22917 used |
| 32 | 161 | 9.6 | 583 | 22 | ABA60344 | Human foetal liver |
| 33 | 161 | 9.6 | 583 | 22 | ABA28591 | Probe #7057 for ge |
| 34 | 161 | 9.6 | 583 | 22 | AAK08624 | Human brain expres |
| 35 | 161 | 9.6 | 583 | 22 | AAK34508 | Human bone marrow |
| 36 | 161 | 9.6 | 583 | 22 | AA116752 | Probe #6685 for ge |
| 37 | 161 | 9.6 | 583 | 22 | AA140226 | Probe #8912 used t |
| 38 | 156 | 9.3 | 156 | 22 | ABA72883 | Human foetal liver |
| 39 | 156 | 9.3 | 156 | 22 | ABA38468 | Probe #16934 for g |
| 40 | 156 | 9.3 | 156 | 22 | AAK21316 | Human brain expres |
| 41 | 156 | 9.3 | 156 | 22 | AAK47474 | Human bone marrow |
| 42 | 156 | 9.3 | 156 | 22 | AA125947 | Probe #15880 for g |
| 43 | 156 | 9.3 | 156 | 22 | AA153305 | Probe #21991 used |
| 44 | 150 | 8.9 | 576 | 21 | AA274887 | Human OREX ORF442 |
| 45 | 141 | 8.4 | 375 | 23 | AA572706 | DNA encoding novel |

ALIGNMENTS

RESULT 1
AA229724
ID AA229724 standard; DNA; 1680 BP.
AC AA229724;
DT 27-MAR-2000 (first entry)
XX
DE Human Lung specific gene-1.
KW Lung Specific Gene; LSG; human; diagnostic marker;
KW prognosticate; Lung cancer; diagnosis; ds.
XX
OS Homo sapiens.
XX
PN WO960160-A1.
XX
PD 25-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10344.
XX
PR 21-MAY-1998; 98US-0086212.
XX
PA (DIAD-) DIADEXUS LLC.
XX
PI Yang F, Macina RA, Sun Y;
XX
DR WPI; 2000-116320/10.
XX
PT A new method for diagnosing, monitoring and staging lung cancer
XX
PS Example 1; Pages 33-34; 40pp; English.

XX The present sequence is a lung specific gene (LSG) from human
CC clone ID 2798946. The LSG has high level of tissue specificity for lungs
CC and is overexpressed in cancerous tissues. The sequence serves as a
CC diagnostic marker for detecting, monitoring, staging and pronosticating
CC lung cancer. The diagnosis involves comparing levels of LSG in samples
CC obtained from patient and normal control.

XX Sequence 1680 BP; 380 A; 503 C; 429 G; 367 T; 1 other;

Query Match 100.0%; Score 1680; DB 21; Length 1680;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1680; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ggtgtcagagataaagtttgactccagaccactgcccgggaagagatgagcgg 60
Db 1 ggtgtcagagataaagtttgactccagaccactgcccgggaagagatgagcgg 60
QY 61 ccgaagactccagctgcccaggtctggaatcctgacttgctccctctgacactgg 120
Db 61 ccgaagactccagctgcccaggtctggaatcctgacttgctccctctgacactgg 120
QY 121 aagaatggccggccgtggagactccactctctgtgtgtgtgtgtgtgtgtgtgt 180
Db 121 aagaatggccggccgtggagactccactctctgtgtgtgtgtgtgtgtgtgtgt 180
QY 181 caagccacctcaagctccactgcaattctcatcctctggcccaaaagtcaagaagaa 240
Db 181 caagccacctcaagctccactgcaattctcatcctctggcccaaaagtcaagaagaa 240
QY 241 ctgacacagagactgtaagaagacacacacgacacacacacacacacacacacac 300
Db 241 ctgacacagagactgtaagaagacacacacacacacacacacacacacacacac 300
QY 301 agtgcacatgctgggaaagacagccgagagatccctgtgtgtgtgtgtgtgtgt 360
Db 301 agtgcacatgctgggaaagacagccgagagatccctgtgtgtgtgtgtgtgtgt 360
QY 361 tecttaagacacatcatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
Db 361 tecttaagacacatcatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
QY 421 agcctcgccacatgacacagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 421 agcctcgccacatgacacagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
QY 481 tcaaacagccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540
Db 481 tcaaacagccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540
QY 541 ccatccgcatggaacacagctgcaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
Db 541 ccatccgcatggaacacagctgcaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
QY 601 ccagccatggagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 660
Db 601 ccagccatggagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 660
QY 661 tagctaaagcagatgataaacctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
Db 661 tagctaaagcagatgataaacctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
QY 721 tgtgtccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
Db 721 tgtgtccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
QY 781 aggtgcccatttccctcagatgtgacgtgtgtgtgtgtgtgtgtgtgtgtgtgt 840
Db 781 aggtgcccatttccctcagatgtgacgtgtgtgtgtgtgtgtgtgtgtgtgtgt 840
QY 841 aggtgtacacattcagcttacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900
Db 841 aggtgtacacattcagcttacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900
```

```
Db 841 aggtgtacacattcagcttacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900
QY 901 ccaatgtgttaataactctgagcttccctggaataatgcccacactgtgacaatcc 960
Db 901 ccaatgtgttaataactctgagcttccctggaataatgcccacactgtgacaatcc 960
QY 961 tcaagcctatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
Db 961 tcaagcctatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
QY 1021 aagaattatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
Db 1021 aagaattatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
QY 1081 gcatcggtctgataatgaaagctgacagaataagctggagatctaccagatctgaaga 1140
Db 1081 gcatcggtctgataatgaaagctgacagaataagctggagatctaccagatctgaaga 1140
QY 1141 tccctaactcaggaacactcccgagtttttatagaccaaagccatgccaaggtggcccaac 1200
Db 1141 tccctaactcaggaacactcccgagtttttatagaccaaagccatgccaaggtggcccaac 1200
QY 1201 tgatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
Db 1201 tgatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
QY 1261 tctgaagccagctcgggaagctcagttttaacacaaaggtgacacactataactaactga 1320
Db 1261 tctgaagccagctcgggaagctcagttttaacacaaaggtgacacactataactaactga 1320
QY 1321 ataacacacagctcgtatcagatccagctgataatgagctgtgtgtgtgtgtgtgtgt 1380
Db 1321 ataacacacagctcgtatcagatccagctgataatgagctgtgtgtgtgtgtgtgtgt 1380
QY 1381 atgttctgaaacatcatcactgagatcatcactccaatcctgtgtgtgtgtgtgtgt 1440
Db 1381 atgttctgaaacatcatcactgagatcatcactccaatcctgtgtgtgtgtgtgtgt 1440
QY 1441 gcaaatatgaatctgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500
Db 1441 gcaaatatgaatctgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500
QY 1501 agtccctcagacacagagatgacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1560
Db 1501 agtccctcagacacagagatgacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1560
QY 1561 ctccctgtcccaagtgagactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1620
Db 1561 ctccctgtcccaagtgagactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1620
QY 1621 gagtatgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1680
Db 1621 gagtatgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1680
```

RESULT 2
ID AAH77948 standard; DNA; 1680 BP.
XX
AC AAH77948;
XX
DT 13-NOV-2001 (first entry)
DE Nucleotide sequence of a human lng103 polypeptide.
KW Human; lung cancer specific gene; LSG; lng103; lung cancer; ss.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 124..1577
FT /*tag= a
FT /product= "lng103"

FT /transl_except= "(pos: 328..329, aa: Gly)"
FT /transl_except= "(pos: 372..374, aa: Val)"
PN WO200161055-AZ.
XX 23-AUG-2001.
XX 20-FEB-2001; 2001WO-US05674.
XX 17-FEB-2000; 2000US-0183188.
XX (DIAD-) DIADEXUS INC.
XX Chen S, Sun Y, Macina RA;
XX MPI: 2001-529917/58.
DR P-PSDB: AAG63976.
XX
PT New lung cancer specific gene for the treatment and diagnosis of lung
PT cancer -
PS
XX
XX Claim 1; Page 112-113; 119pp; English.
CC The present sequence represents a human lung cancer specific gene
CC (LSG), and encodes a polypeptide designated lng103. LSGs are useful
CC in the treatment and diagnosis of lung cancer. The treatment of lung
CC cancer comprises the administration of a molecule which down regulates
CC the expression of an LSG. An immune response can be mounted against a
CC target cell expressing an LSG. Identification of potential therapeutic
CC agents for use in imaging and treating lung cancer which comprises
CC screening molecules for an ability to bind to or decrease expression
CC of an LSG relative to LSG in the absence of the agent where the ability
CC of a molecule to bind to the LSG or decrease expression of the LSG is
CC indicative of the molecule being useful in imaging and treating lung
CC cancer.
XX
XX
SQ Sequence 1680 BP; 380 A; 503 C; 429 G; 367 T; 1 other;

Query Match 96.9%; Score 1628; DB 22; Length 1680;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 ggaagcgccgagagcctcaagcgtgccagcgtcgtgcatcctgcaactgtgcctctga 112
DB 53 ggaagcgccgagagcctcaagcgtgccagcgtcgtgcatcctgcaactgtgcctctga 112
QY 113 caactgggaagatggccgagccgtggaccttcacccctctctgtgtgtgtgagagca 172
DB 113 caactgggaagatggccgagccgtggaccttcacccctctctgtgtgtgtgagagca 172
QY 173 ccttatacaagccacccctcaagtcacacgttcctatacctctggcccaaaagtcata 232
DB 173 ccttatacaagccacccctcaagtcacacgttcctatacctctggcccaaaagtcata 232
QY 233 aagaagaagctgacagagagctgaaagacacacagccacagatcctctgcagagctgc 292
DB 233 aagaagaagctgacagagagctgaaagacacacagccacagatcctctgcagagctgc 292
QY 293 cgtctcagtgcatgctggggaagcaagccgagagatccctgtgtgtgagagctggt 352
DB 293 cgtctcagtgcatgctggggaagcaagccgagagatccctgtgtgtgagagctggt 352
QY 353 gaaacccgtctggaagacatcatctgtgtgaagtcatacagcttaacatctctcagct 412
DB 353 gaaacccgtctggaagacatcatctgtgtgaagtcatacagcttaacatctctcagct 412
QY 413 gaagtgtaagccctgagcaatgacagagagctgtctagtcaaatgcccttgatagct 472
DB 413 gaagtgtaagccctgagcaatgacagagagctgtctagtcaaatgcccttgatagct 472
QY 473 ggtctgattcaacagcccttgctcaagaacatctgtgtgagtcacatgtacgtgaagtc 532
DB 473 ggtctgattcaacagcccttgctcaagaacatctgtgtgagtcacatgtacgtgaagtc 532

DB 473 ggtctgattcaacagcccttgctcaagaacatctgtgtgagtcacatgtacgtgaagtc 532
QY 533 ccaagccacacatccagcatgagacacacagctgagccacccgctgtctctcagta 592
DB 533 ccaagccacacatccagcatgagacacacagctgagccacccgctgtctctcagta 592
QY 593 ctgtgtccacacagcatgagagccctgtgcataccaactgtgtcaaaagctctcctgtgt 652
DB 593 ctgtgtccacacagcatgagagccctgtgcataccaactgtgtcaaaagctctcctgtgt 652
QY 653 gaagccttaagcttaagcagatcatgaacctcctagtgtccatccctgtgccaatctagta 712
DB 653 gaagccttaagcttaagcagatcatgaacctcctagtgtccatccctgtgccaatctagta 712
QY 713 aaacagagctgtccgtgtatcgaggtcctctcaatggcatgtatgcagactctctga 772
DB 713 aaacagagctgtccgtgtatcgaggtcctctcaatggcatgtatgcagactctctga 772
QY 773 gctgtgtgaagctgcccattccctcagcattgacgcgtcgtgagattgacctctgtatcc 832
DB 773 gctgtgtgaagctgcccattccctcagcattgacgcgtcgtgagattgacctctgtatcc 832
QY 833 tgcatacaagggctgacacacatccatcagctcactgtgagcgaagtgtgtgactacaagg 892
DB 833 tgcatacaagggctgacacacatccatcagctcactgtgagcgaagtgtgtgactacaagg 892
QY 893 aaaggtgacacaggtgttataactctgcaacttcctccgaataatggccacactgtgacaa 952
DB 893 aaaggtgacacaggtgttataactctgcaacttcctccgaataatggccacactgtgacaa 952
QY 953 cctccggttcaacccatcgtgtatgacagctgtgtgaagctgagctgtgcgtgtgt 1012
DB 953 cctccggttcaacccatcgtgtatgacagctgtgtgaagctgagctgtgcgtgtgt 1012
QY 1013 ctctccagaagaatcatcatgtgtcctgttgaactctgtctctcgtgagagtgccatcgct 1072
DB 1013 ctctccagaagaatcatcatgtgtcctgttgaactctgtctctcgtgagagtgccatcgct 1072
QY 1073 gaagtcacagcatcgagctgtatgaatgaagaagctgcaatgagctgggataccacagat 1132
DB 1073 gaagtcacagcatcgagctgtatgaatgaagaagctgcaatgagctgggataccacagat 1132
QY 1133 cgtgaagatcctcaactcaagacacatcccgagcttttataagacaaagcgacatgccaagt 1192
DB 1133 cgtgaagatcctcaactcaagacacatcccgagcttttataagacaaagcgacatgccaagt 1192
QY 1193 ggcacacatgatactgtgtggaagtgcttccctccagtgaaagccctccgcttgttca 1252
DB 1193 ggcacacatgatactgtgtggaagtgcttccctccagtgaaagccctccgcttgttca 1252
QY 1253 ccttgagcatcgaaagcagctgggaagctcagtttatacacaagaagtgacacattatatt 1312
DB 1253 ccttgagcatcgaaagcagctgggaagctcagtttatacacaagaagtgacacattatatt 1312
QY 1313 caacttgaataacatcagctgtgacatccagctgtgaatcgaactcgtgagctgtgt 1372
DB 1313 caacttgaataacatcagctgtgacatccagctgtgaatcgaactcgtgagctgtgt 1372
QY 1373 ccaacctgattctctgaaataacatcacaactgtagatcatcactccatctctgtccgaa 1432
DB 1373 ccaacctgattctctgaaataacatcacaactgtagatcatcactccatctctgtccgaa 1432
QY 1433 ccagaatggcaaatgaagctcggggtcccaatgtcatgtgtggaagcccttgagattcga 1492
DB 1433 ccagaatggcaaatgaagctcggggtcccaatgtcatgtgtggaagcccttgagattcga 1492
QY 1493 ggcagctgagctcctcaacttacaagaatgccccttgcttactcaagccctctctgtgaa 1552
DB 1493 ggcagctgagctcctcaacttacaagaatgccccttgcttactcaagccctctctgtgaa 1552
QY 1553 acccagctctcctgtctctccagtgaaagacttgatgtgacacatcaaggaaagctgtgtc 1612
DB 1553 acccagctctcctgtctctccagtgaaagacttgatgtgacacatcaaggaaagctgtgtc 1612

Db 901 accaagtggtcaataactctgcagcttccctgcagatgccacactggaacaacatccg 960
QY 960 tttagagctatcgttgtagtagacgtggtgaaagctgcagctggtctgtctcttcca 1019
Db 961 tttagagctatcgttgtagtagacgtggtgaaagctgcagctggtctgtctcttcca 1020
QY 1020 gaagaatcatgtctctgttgtagctctgtctctcctgaaagtgcacatgcgtgaagta 1079
Db 1021 gaagaatcatgtctctgttgtagctctgtctctcctgaaagtgcacatgcgtgaagta 1080
QY 1080 agcatcgggctgatacaatgaaaggtctgagataagcttgtagtcttaaccagatcgtgaag 1139
Db 1081 agcatcgggctgatacaatgaaaggtctgagataagcttgtagtcttaaccagatcgtgaag 1140
QY 1140 atccctaactcagagacatccctccaggtttttatagaccagaggccaatgcgaagtgcacca 1199
Db 1141 atccctaactcagagacatccctccaggtttttatagaccagaggccaatgcgaagtgcacca 1200
QY 1200 ctgatactgctgaaagtgttccctccagtgaaagccctccgacctgtgttcaacctgggc 1259
Db 1201 ctgatactgctgaaagtgttccctccagtgaaagccctccgacctgtgttcaacctgggc 1260
QY 1260 atcgaagcagctcggaaagctcaagttttacaccaaaagtgcaccaacttatactcaacttg 1319
Db 1261 atcgaagcagctcggaaagctcaagttttacaccaaaagtgcaccaacttatactcaacttg 1320
QY 1320 aataaatacagctctgatactggtaccagctgatactcggatctggatctgctgttccaact 1379
Db 1321 aataaatacagctctgatactggtaccagctgatactcggatctggatctgctgttccaact 1380
QY 1380 gatgtctgaaaaacaacatcacactgagatcacatccatccctgtctgcgaaagcaat 1439
Db 1381 gatgtctgaaaaacaacatcacactgagatcacatccatccctgtctgcgaaagcaat 1440
QY 1440 ggcacaaataagatctgggggtcccaagtgtcatctgttgtaagggccttggagatcgaagcagt 1499
Db 1441 ggcacaaataagatctgggggtcccaagtgtcatctgttgtaagggccttggagatcgaagcagt 1500
QY 1500 gagtccctaactgacaaagagatgccttgtgttactccagcctccttctgttggaaccacac 1559
Db 1501 gagtccctaactgacaaagagatgccttgtgttactccagcctccttctgttggaaccacac 1560
QY 1560 tctccgtctcccaagtgaagacttgtagtgcaagcacaatcagggaaggtctgggtccagctg 1619
Db 1561 tctccgtctcccaagtgaagacttgtagtgcaagcacaatcagggaaggtctgggtccagctg 1620
QY 1620 gagagtaggggtgtagagctctatagaccatccctctctgcaatcaataaacaacttgctgt 1679
Db 1621 gagagtaggggtgtagagctctatagaccatccctctctgcaatcaataaacaacttgctgt 1680
QY 1680 g 1680
Db 1681 g 1681

RESULT 4

AAA37057
ID AAA37057 standard; cDNA; 1636 BP.

AC AAA37057;

DT 08-AUG-2000 (first entry)

DE Human PRO1357 (UNQ706) cDNA sequence SEQ ID NO:127.

XX Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW Transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
ss:

OS Homo sapiens.

XX
XX
PN WO200012708-A2.
XX

PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102309.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.

07-OCT-1998; 98US-0103396.
 PR 07-OCT-1998; 98US-0103401.
 PR 08-OCT-1998; 98US-0103633.
 PR 08-OCT-1998; 98US-0103678.
 PR 08-OCT-1998; 98US-0103679.
 PR 08-OCT-1998; 98US-0103711.
 PR 14-OCT-1998; 98US-0104257.
 PR 20-OCT-1998; 98US-0104987.
 PR 20-OCT-1998; 98US-0105000.
 PR 20-OCT-1998; 98US-0105002.
 PR 21-OCT-1998; 98US-0105104.
 PR 22-OCT-1998; 98US-0105169.
 PR 22-OCT-1998; 98US-0105266.
 PR 26-OCT-1998; 98US-0105693.
 PR 26-OCT-1998; 98US-0105694.
 PR 27-OCT-1998; 98US-0105807.
 PR 27-OCT-1998; 98US-0105881.
 PR 27-OCT-1998; 98US-0105882.
 PR 27-OCT-1998; 98US-0106062.
 PR 28-OCT-1998; 98US-0106023.
 PR 28-OCT-1998; 98US-0106029.
 PR 28-OCT-1998; 98US-0106030.
 PR 28-OCT-1998; 98US-0106032.
 PR 28-OCT-1998; 98US-0106033.
 PR 28-OCT-1998; 98US-0106178.
 PR 29-OCT-1998; 98US-0106248.
 PR 29-OCT-1998; 98US-0106384.
 PR 29-OCT-1998; 98US-0108500.
 PR 30-OCT-1998; 98US-0106464.
 PR 03-NOV-1998; 98US-0106856.
 PR 03-NOV-1998; 98US-0106902.
 PR 03-NOV-1998; 98US-0106905.
 PR 03-NOV-1998; 98US-0106919.
 PR 03-NOV-1998; 98US-0106932.
 PR 10-NOV-1998; 98US-0106934.
 PR 10-NOV-1998; 98US-0107783.
 PR 17-NOV-1998; 98US-0108775.
 PR 17-NOV-1998; 98US-0108779.
 PR 17-NOV-1998; 98US-0108787.
 PR 17-NOV-1998; 98US-0108787.
 PR 17-NOV-1998; 98US-0108801.
 PR 17-NOV-1998; 98US-0108802.
 PR 17-NOV-1998; 98US-0108806.
 PR 17-NOV-1998; 98US-0108807.
 PR 17-NOV-1998; 98US-0108867.
 PR 17-NOV-1998; 98US-0108925.
 PR 18-NOV-1998; 98US-0108848.
 PR 18-NOV-1998; 98US-0108849.
 PR 18-NOV-1998; 98US-0108850.
 PR 18-NOV-1998; 98US-0108851.
 PR 18-NOV-1998; 98US-0108852.
 PR 18-NOV-1998; 98US-0108858.
 PR 18-NOV-1998; 98US-0108904.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
 XX WPI: 2000-237871/20.
 XX P-PSDB: AAY99375.
 DR
 XX
 PT New mammalian DNA sequences encoding transmembrane, receptor or
 PT secreted PRO polypeptides, useful for screening of potential peptide or
 PT small molecule inhibitors of the relevant receptor/Ligand interactions
 XX
 XX Claim 2: Fig 71; 773pp; English.
 CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
 CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
 CC transmembrane and receptor PRO proteins can be used for screening of
 CC potential peptide or small molecule inhibitors of the relevant
 CC receptor/Ligand interactions. The polypeptides and nucleotide sequences
 CC encoding then have various industrial applications, including uses as

CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
 CC PCR primers and hybridisation probes used in the isolation of the PRO
 CC polypeptides from the present invention.
 XX
 SQ Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other;
 Query Match 86.7%; Score 1457; DB 21; Length 1636;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1627; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 53 ggaagcgccgagagccacccagcggcccaagctcgtgacatccctcgtcgcctcga 112
 Db 3 ggaagcgccgagagccacccagcggcccaagctcgtgacatccctcgtcgcctcga 62
 QY 113 cacttggaaagatggccgccttgagccttcacccctcctctgtgttgcggagca 172
 Db 63 cacttggaaagatggccgccttgagccttcacccctcctctgtgttgcggagca 122
 QY 173 ccttgatccaagccacccctagtcacatgctcattcattcctcggcccaagatcata 232
 Db 123 ccttgatccaagccacccctagtcacatgctcattcattcctcggcccaagatcata 182
 QY 233 aagaagaagctgacacaggaagctgaagagccacacacccacacacatccctgcagcagtc 292
 Db 183 aagaagaagctgacacaggaagctgaagagccacacacccacacacatccctgcagcagtc 242
 QY 293 cgcctgctcagtcacatgcgggaaagccagccggaagc-ctccctgctgcggcagcctgg 351
 Db 243 cgcctgctcagtcacatgcgggaaagccagccggaagcctccctgctgcggcagcctgg 302
 QY 352 tgaacacgcgtccctggaagcattatctgtcgaagtgatcaacagatcaacatccctcagc 411
 Db 303 tgaacacgcgtccctggaagcattatctgtcgaagtgatcaacagatcaacatccctcagc 362
 QY 412 tgcagg tgaagccctgcgcgaatgacacagagctgtaagtcaggctccctgcggacatgg 471
 Db 363 tgcagg tgaagccctgcgcgaatgacacagagctgtaagtcaggctccctgcggacatgg 422
 QY 472 tggctgattcaacacgcccctggtcgaagaacatcgtgtgagttcacat tgcagatgagg 531
 Db 423 tggctgattcaacacgcccctggtcgaagaacatcgtgtgagttcacat tgcagatgagg 482
 QY 532 ccacaagccacacatccgcacatgagacacagtgacagggcccaaccccgctgtctcagtg 591
 Db 483 ccacaagccacacatccgcacatgagacacagtgacagggcccaaccccgctgtctcagtg 542
 QY 592 actgtgccaacagcattggagcctgcgcatccaactgtcgtataagcttcctcctcgg 651
 Db 543 actgtgccaacagcattggagcctgcgcatccaactgtcgtataagcttcctcctcgg 602
 QY 652 tgaagccttaagcattgaagctgacacacccctcagtgccatccctcgaattagtga 711
 Db 603 tgaagccttaagcattgaagctgacacacccctcagtgccatccctcgaattagtga 662
 QY 712 aaacacagctgtcccgctgtagcaggtcttcctcaatgagatgatagcagaactctgc 771
 Db 663 aaacacagctgtcccgctgtagcaggtcttcctcaatgagatgatagcagaactctgc 722
 QY 772 agctgtgtaaggtgccattccctcagcattgacagctgtgagattgacattctgatac 831
 Db 723 agctgtgtaaggtgccattccctcagcattgacagctgtgagattgacattctgatac 782
 QY 832 ctgcacatcaaggtgacacacattcagcttaccctgggggccaaggtgttgactcaacagg 891
 Db 783 ctgcacatcaaggtgacacacattcagcttaccctgggggccaaggtgttgactcaacagg 842
 QY 892 gaaagtgacacagtggttcaataactcgtcagcttcccggaacatgcccacccctgaca 951
 Db 843 gaaagtgacacagtggttcaataactcgtcagcttcccggaacatgcccacccctgaca 902
 QY 952 acatcccggttcagcctcatcgtgagtcaggaagctgtgaaagctgagctgctgtgctg 1011

Db 903 acatccgcttcagcctcatcgtgagtcagagcgtggtgaagctgacgtgctgtgtgc 962
 Qy 1012 tctccagaagaattcgtgctcgtgttgactctgtcttcttaagagtgccctggc 1071
 Db 963 tctccagaagaattcagagtcctgtgtgactctgtcttcttaagagtgccctggc 1022
 Qy 1072 tgaagtcagagcctggcgtgactcaatgaagagcgtcagatgaagtgatctaccaga 1131
 Db 1023 tgaagtcagagcctggcgtgactcaatgaagagcgtcagatgaagtgatctaccaga 1082
 Qy 1132 tctgtaagatccttaactaagagcactccgagtttttataagaccagagccatgcaga 1191
 Db 1083 tctgtaagatccttaactaagagcactccgagtttttataagaccagagccatgcaga 1142
 Qy 1192 tggccaactgactgctgtgagagtggttccctcagatgaagccttcgcccctgtgtca 1251
 Db 1143 tggccaactgactgctgtgagagtggttccctcagatgaagccttcgcccctgtgtca 1202
 Qy 1252 cccctggagtcgaagcagcagctcgaagctcagtttatacaccagagtgaccaactatac 1311
 Db 1203 cccctggagtcgaagcagcagctcgaagctcagtttatacaccagagtgaccaactatac 1262
 Qy 1312 tcaacttgaataacatcagcctctgtatcggatccagctgatactctggagtgctggt 1371
 Db 1263 tcaacttgaataacatcagcctctgtatcggatccagctgatactctggagtgctggt 1322
 Qy 1372 tccaactgagtgcttctgaagaaacatcatcagcagatatacctaccatccctgcgcga 1431
 Db 1323 tccaactgagtgcttctgaagaaacatcatcagcagatatacctaccatccctgcgcga 1382
 Qy 1432 accagaatgagcaaatlaagatcttggtgtccagatgctatctgtgaagccttgagatcg 1491
 Db 1383 accagaatgagcaaatlaagatcttggtgtccagatgctatctgtgaagccttgagatcg 1442
 Qy 1492 aggcagcgtgagtcctcactgacagagagatgcccctgtgtactactcagcctctgtgga 1551
 Db 1443 aggcagcgtgagtcctcactgacagagagatgcccctgtgtactactcagcctctgtgga 1502
 Qy 1552 aaccagagcctcctgtctccccaagatgaagctgagtgagagcagcagagagagctggt 1611
 Db 1503 aaccagagcctcctgtctccccaagatgaagctgagtgagagcagcagagagagctggt 1562
 Qy 1612 cccaagtcggagtgagtggtgtgagctctatagacatccctctctgcaatacaataacac 1671
 Db 1563 cccaagtcggagtgagtggtgtgagctctatagacatccctctctgcaatacaataacac 1622
 Qy 1672 ttgcctgtg 1680
 Db 1623 ttgcctgtg 1631

RESULT 5
 AAS46064
 ID AAS46064 standard; cDNA; 1636 BP.
 AC AAS46064;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human DNA encoding PRO polypeptide sequence #140.
 XX
 KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
 KW PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN WO200168848-A2.
 XX
 PD 20-SEP-2001.
 XX

PF 28-FEB-2001; 2001WO-US06520.
 XX
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 03-MAR-2000; 2000US-187202P.
 PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 14-MAR-2000; 2000US-189328P.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 29-MAR-2000; 2000US-193032P.
 PR 29-MAR-2000; 2000US-193053P.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 04-APR-2000; 2000US-194449P.
 PR 04-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000WO-US14705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-064484P.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 PA (GENTH) GENENTECH INC.
 XX
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI; 2001-602746/68.
 DR P-PSDB; AAU29163.
 XX
 PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 Claim 2: Fig 279; 774pp; English.

Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
 primers for PRO polypeptides of the invention. The sequences of the
 CC invention can be used to detect the presence of a tumour in a mammal by
 CC comparing the level of expression of a PRO polypeptide in a test sample
 CC of cells from the animal and a control sample of normal cells, whereby a
 CC higher level of expression in the test sample indicates the presence of a
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
 CC pigs, goats and rabbits but are preferably human. The polypeptides can be
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human
 CC blood, when contacted with it. A specific polypeptide can be used to
 CC stimulate the proliferation or differentiation of chondrocyte cells. The
 CC PRO proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.

Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other;

Query Match 86.7%; Score 1457; DB 22; Length 1636;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1627; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
QY 53 ggaagcgagccagagatccagcgtgcccagctgcatccctgcaactgttcctctga 112
Db 3 ggaagcgagccagagatccagcgtgcccagctgcatccctgcaactgttcctctga 62
QY 113 caacttggaagatgagccggtccgttggaaccttaacctctgtgtgtctgagca 172
Db 63 caacttggaagatgagccggtccgttggaaccttaacctctgtgtgtctgagca 122
QY 173 ccttgatccaagccacccctcagctcccaactgcaattctatctctgcccacaaagtca 232
Db 123 ccttgatccaagccacccctcagctcccaactgcaattctatctctgcccacaaagtca 182
QY 233 aagaaaagctgacacagagagctgaaggaacaaacgccaacagacatccctgacagctgc 292
Db 183 aagaaaagctgacacagagagctgaaggaacaaacgccaacagacatccctgacagctgc 242
QY 293 cgcctctcagctgacatgctgaggaagacagccagcgagag- atccctgtctggtgagcagctg 351
Db 243 cgcctctcagctgacatgctgaggaagacagccagcgagagcctcctgtgtgtgagcagctg 302
QY 352 tgaacacgcgtctgacacatcatctgctgaaggtcatcacagctacaactccctcagc 411
Db 303 tgaacacgcgtctgacacatcatctgctgaaggtcatcacagctacaactccctcagc 362
QY 412 tgaaggttaagccctcgcgcacatgacagagagctgctagcttaagaatccctctgacatg 471
Db 363 tgcaggtgaagcctcctgcgcacatgacagagagctgctagcttaagaatccctctgacatg 422
QY 472 tgcgtgattcaacacgcccctgtgtcaagaacatcgttgaggttccacatgacagctag 531
Db 423 tgcgtgattcaacacgcccctgtgtcaagaacatcgttgaggttccacatgacagctag 482
QY 532 cccaagccacatccgcatgtagacacacagctgaagtggcccacccgcctgtgctccatg 591
Db 483 cccaagccacatccgcatgtagacacacagctgaagtggcccacccgcctgtgctccatg 542
QY 592 actgtgcacacagacagagagcctgctgcatccaaactgctgcatatagcttctctctg 651
Db 543 actgtgcacacagacagagagcctgctgcatccaaactgctgcatatagcttctctctg 602
QY 652 tgaacgccttaagctaaagcaggtcatgaaacctcctagtccatccctgcccacatcagta 711
Db 603 tgaacgccttaagctaaagcaggtcatgaaacctcctagtccatccctgcccacatcagta 662
QY 712 aaaaacagcctgtcccgtagtcgagcttccttaagctcgtatgacatgacacctctgc 771
Db 663 aaaaacagcctgtcccgtagtcgagcttccttaagctcgtatgacatgacacctctgc 722
QY 772 agctgtgtgaaggtgcccatttccctcagaatctgacagctgtgagtttgacactctg 831
Db 723 agctgtgtgaaggtgcccatttccctcagaatctgacagctgtgagtttgacactctg 782
QY 832 ctgcatcaaggtgtgacacatctacgttacctggtgggcccagatgttggactcaag 891
Db 783 ctgcatcaaggtgtgacacatctacgttacctggtgggcccagatgttggactcaag 842
QY 892 gaaaggtgaccaaagtgttcaataactctgacagcttccctgccaatgcccacccgga 951
Db 843 gaaaggtgaccaaagtgttcaataactctgacagcttccctgccaatgcccacccgga 902
QY 952 acatccgcttcagccctcagctgagtcagagacgttgtaagctcgaatgagctgtctg 1011
Db 903 acatccgcttcagccctcagctgagtcagagacgttgtaagctcgaatgagctgtctg 962
QY 1012 tctctccagaagaatcagtgccctgtgagctgtcttctcctaaggtgccatcggc 1071
Db 1071 tctctccagaagaatcagtgccctgtgagctgtcttctcctaaggtgccatcggc 1022
```

```
Db 963 tctctccagaagaatcagtgccctgtgagctgtcttctcctaaggtgccatcggc 1022
QY 1072 tgaagtcagaagctgggctgatacctaataaaggctgtaagatagctgggataccaga 1131
Db 1023 tgaagtcagaagctgggctgatacctaataaaggctgtaagatagctgggataccaga 1082
QY 1132 tcgtgaagatcctaactcagagacactcccgagtttttatacagaacgaagcctgcccag 1191
Db 1083 tcgtgaagatcctaactcagagacactcccgagtttttatacagaacgaagcctgcccag 1142
QY 1192 tggcccaactgatacgtgtgtggaagtggttccctccagttgaagccctccgcttgttca 1251
Db 1143 tggcccaactgatacgtgtgtggaagtggttccctccagttgaagccctccgcttgttca 1202
QY 1252 cccctgggcatcgaagccagctcggagacatgatttacaacaaaggctgccaattatc 1311
Db 1203 cccctgggcatcgaagccagctcggagacatgatttacaacaaaggctgccaattatc 1262
QY 1312 tcaacttgaataacatcagctgatacagatccagctga tgaactctggatgtgctggt 1371
Db 1263 tcaacttgaataacatcagctgatacagatccagctga tgaactctggatgtgctggt 1322
QY 1372 tccaactgattgtctgtaaaaacatcactgaactgaatccatccatccctgctgcga 1431
Db 1323 tccaactgattgtctgtaaaaacatcactgaactgaatccatccatccctgctgcga 1382
QY 1432 accagaatggcacaattaaagatctgggtccagtgatctgtgtgaaggtccttgatctg 1491
Db 1383 accagaatggcacaattaaagatctgggtccagtgatctgtgtgaaggtccttgatctg 1442
QY 1492 aggcagctgagctcctcactcagacagagatgacctgtgttacttccacactcctgtgga 1551
Db 1443 aggcagctgagctcctcactcagacagagatgacctgtgttacttccacactcctgtgga 1502
QY 1552 aacccagctctctgtctccagtgaaagacttga tggcagccatcagaaggagctggt 1611
Db 1503 aacccagctctctgtctccagtgaaagacttga tggcagccatcagaaggagctggt 1562
QY 1612 cccaactgggaatgaggtgtgagctctatagacacatccctctctgcaataataaacac 1671
Db 1563 cccaactgggaatgaggtgtgagctctatagacacacatccctctctgcaataataaacac 1622
QY 1672 ttgcctgtg 1680
Db 1623 ttgcctgtg 1631
```

RESULT 6
AAF92096
ID AAF92096 standard; cDNA; 1636 BP.
XX
AC AAF92096;
XX
DT 15-MAY-2001 (first entry)
XX
DE Human PRO1357 cDNA.
XX
KW Human; PRO protein; mapping; ss.
XX
OS Homo sapiens.
XX
PN W0200116318-A2.
XX
PD 08-MAR-2001.
XX
PE 24-AUG-2000; 2000MO-US23328.
XX
PR 01-SEP-1999; 99MO-US20111.
PR 15-SEP-1999; 99MO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000MO-US04341.

PR 18-FEB-2000; 2000MO-US04342.
 PR 22-FEB-2000; 2000MO-US04414.
 PR 01-MAR-2000; 2000MO-US05601.
 PR 03-MAR-2000; 2000MO-US0187202.
 PR 25-APR-2000; 2000MO-US0193997.
 PR 22-MAY-2000; 2000MO-US14042.
 PR 05-JUN-2000; 2000MO-US0209832.
 XX
 XX (GENTH) GENENTECH INC.
 XX
 PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
 PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
 XX
 DR WPI; 2001-183260/18.
 DR P-PSDB; AAB87564.
 XX
 PT Eighty four nucleic acids encoding PRO polypeptides, useful in
 PT molecular biology, including use as hybridization probes, and in
 PT chromosome and gene mapping.
 XX
 PS Claim 2; Fig 77; 278pp; English.
 CC The present sequence is the coding sequence for a human PRO polypeptide
 CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO
 CC antagonists or anti-PRO antibodies are useful for preparation of a
 CC medicament useful in the treatment of a condition which is responsive to
 CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO
 CC protein may also be employed as molecular weight markers for protein
 CC electrophoresis. The PRO coding sequence has applications in molecular
 CC biology, including use as hybridisation probes, and in chromosome and
 CC gene mapping.
 XX
 SO Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other;

Query Match 86.7%; Score 1457; DB 22; Length 1636;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1627; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 53 ggaagcgccgagagatccagcgtgcccaggtctgcatctcgcagcttgcctctga 112
 DB 3 ggaagcgccgagagatccagcgtgcccaggtctgcatctcgcagcttgcctctga 62
 QY 113 caccctggagagatgagcggccctgagacctcaacctctctgtgtgttgctggagca 172
 DB 63 caccctggagagatgagcggccctgagacctcaacctctctgtgtgttgctggagca 122
 QY 173 ccttgatccaagccacctcagctcccaatgcatctcattcctcctggcccaaatgcatca 232
 DB 123 ccttgatccaagccacctcagctcccaatgcatctcattcctcctggcccaaatgcatca 182
 QY 233 aagaaagctgacacagagagctgaagacacaaagcccaacagatctctgcaagctgc 292
 DB 183 aagaaagctgacacagagagctgaagacacaaagcccaacagatctctgcaagctgc 242
 QY 293 cctgttcagatgacatgagcgggaaagccagccggaagc-attcctgtgctgggagctgg 351
 DB 243 cctgttcagatgacatgagcgggaaagccagccggaagc-attcctgtgctgggagctgg 302
 QY 352 tgaacacgcttcgtgaagacatcatctgtgctgaagatcacaagatcaatcctcagc 411
 DB 303 tgaacacgcttcgtgaagacatcatctgtgctgaagatcacaagatcaatcctcagc 362
 QY 412 tgaagtgaaagcctctgcccgaatgacagagagctgctagtcaagaatccctctgacatg 471
 DB 363 tgaagtgaaagcctctgcccgaatgacagagagctgctagtcaagaatccctctgacatg 422
 QY 472 tggctgtattcaaacagcctctgtgtaagaacatgctgagatccacatgacagacgaag 531
 DB 423 tggctgtattcaaacagcctctgtgtaagaacatgctgagatccacatgacagacgaag 482
 QY 532 cccaagccacatccgcatgagacacagtgcaagtggcccccgcctgtcctcagtg 591
 DB 1636 cccaagccacatccgcatgagacacagtgcaagtggcccccgcctgtcctcagtg 1622

DB 483 cccaagccacatccgcatgagacacagtgcaagtggcccccgcctgtcctcagtg 542
 QY 592 actgtgccacccagccatgaggagctctgcgcatccaatctgtgataagctctcctctgg 651
 DB 543 actgtgccacccagccatgaggagctctgcgcatccaatctgtgataagctctcctctgg 602
 QY 652 tgaacgcttaagcagcagatcatgaacctcctagtgcacatccctggccaatctagtga 711
 DB 603 tgaacgcttaagcagcagatcatgaacctcctagtgcacatccctggccaatctagtga 662
 QY 712 aaaaacagctgtgtcccgatgacagagcttccctccaatgagatgataagacatctctgc 771
 DB 663 aaaaacagctgtgtcccgatgacagagcttccctccaatgagatgataagacatctctgc 722
 QY 772 agctgtgtaagtgcccatcttccctcagatgagacagctctgtgagtttgacctgtacc 831
 DB 723 agctgtgtaagtgcccatcttccctcagatgagacagctctgtgagtttgacctgtacc 782
 QY 832 ctggccatcaaggtgacacatcagctcagctgaggggccaagtgtgtgagatccacag 891
 DB 783 ctggccatcaaggtgacacatcagctcagctgaggggccaagtgtgtgagatccacag 842
 QY 892 gaaagtgaccaaagtggttcaataaactctgcagcttccctgacaatgacacccctggaca 951
 DB 843 gaaagtgaccaaagtggttcaataaactctgcagcttccctgacaatgacacccctggaca 902
 QY 952 acatcccgcttcagccatcagtgagatcagaagctggtgaaagctgagtgctgctgtgc 1011
 DB 903 acatcccgcttcagccatcagtgagatcagaagctggtgaaagctgagtgctgctgtgc 962
 QY 1012 tctctcagaagaatcatggtctctgtgagctgtgcttccctgagatgagccatccgcg 1071
 DB 963 tctctcagaagaatcatggtctctgtgagctgtgcttccctgagatgagccatccgcg 1022
 QY 1072 tgaagtcagaagcctcgggctgataatgaataaagctgcagataaagctggagatccacaga 1131
 DB 1023 tgaagtcagaagcctcgggctgataatgaataaagctgcagataaagctggagatccacaga 1082
 QY 1132 tctgtgaagatcccaactcagagacatcccgagtttttataagccaagccatgccaag 1191
 DB 1083 tctgtgaagatcccaactcagagacatcccgagtttttataagccaagccatgccaag 1142
 QY 1192 tggcccaactgacgtgctggaagtggttccctcagatgaagccctccgcttctga 1251
 DB 1143 tggcccaactgacgtgctggaagtggttccctcagatgaagccctccgcttctga 1202
 QY 1252 cctctgggcatcgaaagcagctcgaaagctcaagttttacacccaaggtgaaaccaatatac 1311
 DB 1203 cctctgggcatcgaaagcagctcgaaagctcaagttttacacccaaggtgaaaccaatatac 1262
 QY 1312 tcaacttgaataaatacagctctgatacagatcagatgaactctgagattgctgtgt 1371
 DB 1263 tcaacttgaataaatacagctctgatacagatcagatgaactctgagattgctgtgt 1322
 QY 1372 tccaactctatgtcttgaaanaaacatcatcactagatgatacctcactcactctgtgcga 1431
 DB 1323 tccaactctatgtcttgaaanaaacatcatcactagatgatacctcactcactctgtgcga 1382
 QY 1432 accagaatgagcaaatgaatctgaggtccagtgatcatctgtgaaagccttggatctcg 1491
 DB 1383 accagaatgagcaaatgaatctgaggtccagtgatcatctgtgaaagccttggatctcg 1442
 QY 1492 aggcagctgagctctcactacgacaaagatgacctgtgttactccaagctctctgtgga 1551
 DB 1443 aggcagctgagctctcactacgacaaagatgacctgtgttactccaagctctctgtgga 1502
 QY 1552 aaccagctctcctctgctccagtgaaagcttgatgagcagccatcagggaagctgggt 1611
 DB 1503 aaccagctctcctctgctccagtgaaagcttgatgagcagccatcagggaagctgggt 1562
 QY 1612 cccaagctggagatagtggtgtgagctctatagacatccctctctgcaataaataaac 1671
 DB 1563 cccaagctggagatagtggtgtgagctctatagacatccctctctgcaataaataaac 1622

OY 1672 ttgctgtg 1680
 |||||
 Db 1623 ttgctgtg 1631

RESULT 7

AAFS4293
 ID AAF54293 standard; DNA; 1636 BP.
 AC AAF54293;

DT 02-APR-2001 (first entry)

DE DNA encoding protein of the invention #36.

KW Secreted; transmembrane; gene therapy; ss.

OS Unidentified.

XX WO200078961-A1.

PD 28-DEC-2000.

PF 18-FEB-2000; 2000WO-US04342.

PR 23-JUN-1999; 99US-0141037.

PR 20-JUL-1999; 99US-014758.

PR 26-JUL-1999; 99US-0145698.

PR 01-SEP-1999; 99WO-US20111.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 02-DEC-1999; 99WO-US28551.

PR 16-DEC-1999; 99WO-US30095.

PR 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00376.

XX (GETH) GENENTECH INC.

PA Baker KP, Bolstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

PI Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-071395/08.

DR Secreted and transmembrane proteins and nucleic acids designated PRO,

PT useful as hybridization probes, in chromosome and gene mapping and gene

PT therapy.

XX Claim 2; Fig 71; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.

CC These proteins and the DNA encoding them may be used as hybridization

CC probes, in chromosome and gene mapping and in the generation of

CC anti-sense RNA and DNA. They may also be used to generate either

CC transgenic animals or knockout animals which are in turn useful for

CC development and screening of therapeutically useful reagents.

CC The nucleic acids may also be used in gene therapy.

XX Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other;

SO

Query Match 86.7%; Score 1457; DB 22; Length 1636;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1627; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 53 gggagggcgagagactcagcgtgcccaggtctggcactcctgctgctcctcga 112

Db 3 gggagggcgagagactcagcgtgcccaggtctggcactcctgctgctcctcga 62

OY 113 cacttgggaagatggcgccgctgtgacttaacccctctctgtgtgttctgtgagca 172

Db 113 cacttgggaagatggcgccgctgtgacttaacccctctctgtgtgttctgtgagca 172

Db 63 cacttgggaagatggcgccgctgtgacttaacccctctctgtgtgttctgtgagca 122

OY 173 ccttgatccaagcaccctcagtcaccctgcaatcttcacatcttcctgagcccaagatca 232

Db 123 ccttgatccaagcaccctcagtcaccctgcaatcttcacatcttcctgagcccaagatca 182

OY 233 aagaagaagctgaacagaagctgaagagcacaacagcccaagcactcctgcaagctgc 292

Db 183 aagaagaagctgaacagaagctgaagagcacaacagcccaagcactcctgcaagctgc 242

OY 293 cgtctcagctgacatcagcaggaagcagcagcagcagcagcagcagcagcagcagcagc 351

Db 243 cgtctcagctgacatcagcaggaagcagcagcagcagcagcagcagcagcagcagcagc 302

OY 352 tgaacaccgtctcgaacacatcatctgtgtgaagctgaagctgaagctgaagctgaagc 411

Db 303 tgaacaccgtctcgaacacatcatctgtgtgaagctgaagctgaagctgaagctgaagc 362

OY 412 tgcaggtgaagccctcgtgccaatgaacagagctgtgtgaagctgaagctgaagctgaagc 471

Db 363 tgcaggtgaagccctcgtgccaatgaacagagctgtgtgaagctgaagctgaagctgaagc 422

OY 472 tggctggtatcaaacagccctgtgtcaagacatctgtgtgaagctgaagctgaagctgaagc 531

Db 423 tggctggtatcaaacagccctgtgtcaagacatctgtgtgaagctgaagctgaagctgaagc 482

OY 532 cccaagcacaacatccagatgagacacagctgaagctgaagctgaagctgaagctgaagc 591

Db 483 cccaagcacaacatccagatgagacacagctgaagctgaagctgaagctgaagctgaagc 542

OY 592 actgtgccaccagcacaatggagcctgtgcaacacatctgtgtgaagctgaagctgaagc 651

Db 543 actgtgccaccagcacaatggagcctgtgcaacacatctgtgtgaagctgaagctgaagc 602

OY 652 tgaagccttagctgaagcagctgaacacacacacacacacacacacacacacacacacac 711

Db 603 tgaagccttagctgaagcagctgaacacacacacacacacacacacacacacacacacac 662

OY 712 aaacacagctgtgtccctgagcagcctcctccttaagagctgaagctgaagctgaagc 771

Db 663 aaacacagctgtgtccctgagcagcctcctccttaagagctgaagctgaagctgaagc 722

OY 772 agctgtgaagctgtccacatcttcctcctcctcctcctcctcctcctcctcctcctcctc 831

Db 723 agctgtgaagctgtccacatcttcctcctcctcctcctcctcctcctcctcctcctcctc 782

OY 832 ctgcacatcaagctgaac 891

Db 783 ctgcacatcaagctgaac 842

OY 892 gaaagctgaac 951

Db 843 gaaagctgaac 902

OY 952 acatccgttgaagctgaac 1011

Db 903 acatccgttgaagctgaac 962

OY 1012 tctctcgaagaatctcagctgc 1071

Db 963 tctctcgaagaatctcagctgc 1022

OY 1072 tgaagctgaac 1131

Db 1023 tgaagctgaac 1082

OY 1132 tctgtgaagatcttaac 1191

Db 1083 tctgtgaagatcttaac 1142

OY 1192 tggcccaacatcgtctgtgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1251

Db 1143 tggcccaacatcgtctgtgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1202

| | | | |
|----|------|---|-------|
| QY | 1252 | ccccgggcatcgaagccagctcggaaagctcgaattttaaccsaaagtgaccaactatac | 13111 |
| Db | 1203 | ccccgggcatcgaagccagctcggaaagctcgaattttaaccsaaagtgaccaactatac | 1262 |
| QY | 1312 | tcaacttgaataaatacagctctgcacggatccaagcttgaatgaaccttggaattgctgt | 13717 |
| Db | 1263 | tcaacttgaataaatacagctctgcacggatccaagctctgaactcttggaattgctgt | 13222 |
| QY | 1372 | tccaactctgatttctctgnaaaacatcatcaacttgaatcatccaactcatctctgcgcga | 14313 |
| Db | 1323 | tccaactctgatttctctgnaaaacatcatcaacttgaatcatccaactcatctctgcgcga | 1382 |
| QY | 1432 | accgaatgccaatthaagctctgggtctcccgatgcatcttgttgaagccttggattccg | 1491 |
| Db | 1383 | accgaatgccaatthaagctctgggtctcccgatgcatcttgttgaagccttggattccg | 1442 |
| QY | 1492 | aggcagactgagctctctcaactgacccaagaatgaccttgcttgaatccaagctcctgtgta | 1551 |
| Db | 1443 | aggcagactgagctctctcaactgacccaagaatgaccttgcttgaatccaagctcctgtgta | 1502 |
| QY | 1552 | aaaccaagctctctgtctctccagtgaagacttggatgagcagccatcaagggaagctggct | 16111 |
| Db | 1503 | aaaccaagctctctgtctctccagtgaagacttggatgagcagccatcaagggaagctggct | 1562 |
| QY | 1612 | cccagctcgggaagatggttgcttgaagcctataaacatccctcttcgcataataaacaac | 1671 |
| Db | 1563 | cccagctcgggaagatggttgcttgaagcctataaacatccctcttcgcataataaacaac | 1622 |
| QY | 1672 | ttgcctgtg 1680 | |
| Db | 1623 | ttgcctgtg 1631 | |

| RESULT | 8 |
|----------|---|
| AAH77950 | |
| ID | AAH77950 standard; DNA; 2036 BP. |
| XX | |
| XX | |
| AC | AAH77950; |
| XX | |
| DT | 13-NOV-2001 (first entry) |
| XX | |
| DE | Nucleotide sequence of a human Lng103 polypeptide. |
| XX | |
| KW | Human; lung cancer specific gene; LSG; Lng103; lung cancer; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| FH | Location/Qualifiers |
| FT | 163..1617 |
| CDS | /*tag= a |
| FT | /product= "Lng103" |
| FT | /transl_except= "(pos: 412..414, aa: Val)" |
| XX | |
| PN | WO200161055-A2. |
| XX | |
| PD | 23-AUG-2001. |
| XX | |
| PF | 20-FEB-2001; 2001WO-USO5674. |
| XX | |
| PR | 17-FEB-2000; 2000US-0183188. |
| XX | |
| PA | (DIAD-) DIADEXUS INC. |
| XX | |
| PI | Chen S, Sun Y, Macina RA; |
| XX | |
| DR | WPI; 2001-529917/58. |
| XX | |
| DR | P-PSDB; AAG63976. |
| XX | |
| PT | New lung cancer specific gene for the treatment and diagnosis of lung |
| XX | |
| PS | cancer - |
| XX | |
| PS | Claim 1; Page 116-117; 119pp; English. |

XX The present sequence represents a human lung cancer specific gene
CC (LSG), and encodes a polypeptide designated Lng103. LSGs are useful
CC in the treatment and diagnosis of lung cancer. The treatment of lung
CC cancer comprises the administration of a molecule which down regulates
CC the expression of an LSG. An immune response can be mounted against a
CC target cell expressing an LSG. Identification of potential therapeutic
CC agents for use in imaging and treating lung cancer which comprises
CC screening molecules for an ability to bind to or decrease expression
CC of an LSG relative to LSG in the absence of the agent where the ability
CC of a molecule to bind to the LSG or decrease expression of the LSG is
CC indicative of the molecule being useful in imaging and treating lung
CC cancer.
XX
SQ Sequence 2036 BP; 477 A; 586 C; 528 G; 444 T; 1 other;

| | | | | |
|-----------------------|--------------|--------------|---------------|--------------|
| Query Match | 86.7%; | Score 1457; | DB 22; | Length 2036; |
| Best Local Similarity | 99.9%; | Pred. No. 0; | | |
| Matches 1627; | Conservative | 0; | Mismatches 1; | Indels 1; |
| | | | Gaps | 1; |

| | | | |
|----|-----|---|-----|
| QY | 53 | ggagcggcgccgagagatcccaacgctgagccaaagctctgcatctctgacacttgctgcgcctctga | 112 |
| Db | 92 | ggagcgcgccgagagatcccaacgctgagccaaagctctgcatctctgacacttgctgcgcctctga | 151 |
| QY | 113 | caaccgggaaatgagccggcccgctggagacttcacacctctctgagcttggctggcgaacca | 172 |
| Db | 152 | caaccgggaaatgagccggcccgctggagacttcacacctctctgagcttggctggcgaacca | 211 |
| QY | 173 | ccctgatccaaagccaccctcagctcccaactgacttctatccctctgagcccaaaatctca | 232 |
| Db | 212 | ccctgatccaaagccaccctcagctcccaactgacttctatccctctgagcccaaaatctca | 271 |
| QY | 233 | aagaaagctctgacacagagagcttgaaagacccaacgcgcacacagcatccctgacagctgc | 292 |
| Db | 272 | aagaaagctctgacacagagagcttgaaagacccaacgcgcacacagcatccctgacagagctgc | 331 |
| QY | 293 | cgcctctaaagctgacatggcggaagacagcccggaag-actccctgctctggcagagcttg | 355 |
| Db | 332 | cgcctctaaagctgacatggcggaagacagcccggaagacatccctgctctggcagagcttg | 391 |
| QY | 352 | tgaacaacgctccctgaaagacacatcatctgctcgaaaggtcatatcacagcgcacaaatccctcagc | 411 |
| Db | 392 | tgaacaacgctccctgaaagacacatcatctgctcgaaaggtcatatcacagcgcacaaatccctcagc | 455 |
| QY | 412 | tgcagagtgaaagcctctcggcccaatgacccaagagctgctagatcagaatcccccctggaacatgg | 471 |
| Db | 452 | tgcagagtgaaagcctctcggcccaatgacccaagagctgctagatcagaatcccccctggaacatgg | 511 |
| QY | 472 | tggcctggagcttcaaacgcgcctctgctcgaagacacatcgctggagcttccaatagacagcttagag | 531 |
| Db | 512 | tggcctggagcttcaaacgcgcctctgctcgaagacacatcgctggagcttccaatagacagcttagag | 571 |
| QY | 532 | cccaagcgcacacatcccgcaatggacacaaatgacaagtggccccaccgcgcctgctccacagctg | 591 |
| Db | 572 | cccaagcgcacacatcccgcaatggacacaaatgacaagtggccccaccgcgcctgctccacagctg | 631 |
| QY | 592 | actgtgtccacacagacatgggagcctgtcgatccatcaactgtgtgcataaagctctcctcctgg | 651 |
| Db | 632 | actgtgtccacacagacatgggagcctgtcgatccatcaactgtgtgcataaagctctcctcctgg | 691 |
| QY | 652 | tgaagcctcttgcttaagacagctcatgtaaacctccctcagctgacatccctgtcccaatcttagtga | 711 |
| Db | 692 | tgaagcctcttgcttaagacagctcatgtaaacctccctcagctgacatccctgtcccaatcttagtga | 751 |
| QY | 712 | aaaacacagctgtgtcccgctgaatcagagctcttccctcaatgtgatatgagacacctctgc | 771 |
| Db | 752 | aaaacacagctgtgtcccgctgaatcagagctcttccctcaatgtgatatgagacacctctctgc | 811 |
| QY | 772 | agcttggtgaaagctgcccattctccctcagacatggaacgctctgagcttgagacttgacaccttgatc | 831 |
| Db | 812 | agcttggtgaaagctgcccattctccctcagacatggaacgctctgagacttgagacttgacaccttgatc | 871 |

```
QY 832 ctgcacatcaagggtgacacatcaccatctaccctgggggccaagtgtgtgacacagg 891
   |||||||
Db 872 ctgcacatcaagggtgacacatcaccatctaccctgggggccaagtgtgtgacacagg 931
QY 892 gaaaggtgaccagtggttcaataactctgcagcttccctgacacatgcccactggaca 951
   |||||||
Db 932 gaaaggtgaccagtggttcaataactctgcagcttccctgacacatgcccactggaca 991
QY 952 acatcccgctcagcctcaccatctgagtcagagagctggtaaaagctgagcgtgtctgtgc 1011
   |||||||
Db 992 acatcccgctcagcctcaccatctgagtcagagagctggtaaaagctgagcgtgtctgtgc 1051
QY 1012 tctctcagaagaatcaccatgctcgtgtgtgactctgtgtcttccctgagagtgccatcggc 1071
   |||||||
Db 1052 tctctcagaagaatcaccatgctcgtgtgtgactctgtgtcttccctgagagtgccatcggc 1111
QY 1072 tgaagtcacagcatcggtgcgcatcaccatgtaaaagctgcacataagcttgatctaccaga 1131
   |||||||
Db 1112 tgaagtcacagcatcggtgcgcatcaccatgtaaaagctgcacataagcttgatctaccaga 1171
QY 1132 tctgtaagatctctaactcagcagcactcccgagtttttatagaccagaagcagtcgcaag 1191
   |||||||
Db 1172 tctgtaagatctctaactcagcagcactcccgagtttttatagaccagaagcagtcgcaag 1231
QY 1192 tggcccaactgactcgtctgtagaagtgttccctcagtgaaagccttcgccccttgttca 1251
   |||||||
Db 1232 tggcccaactgactcgtctgtagaagtgttccctcagtgaaagccttcgccccttgttca 1291
QY 1252 ccctgtgcatcgaaagcagcctcggaagctcagtttaccaccaaggtgaccacattatc 1311
   |||||||
Db 1292 ccctgtgcatcgaaagcagcctcggaagctcagtttaccaccaaggtgaccacattatc 1351
QY 1312 tcaactgataataacatcagcctctgtcgtgacagcgtgtatgaactctgggattgtgcgtg 1371
   |||||||
Db 1352 tcaactgataataacatcagcctctgtcgtgacagcgtgtatgaactctgggattgtgcgtg 1411
QY 1372 tccaactgtatgttcttgaaaaaacatcaccatcagatgatcatccatccatctcgtgcga 1431
   |||||||
Db 1412 tccaactgtatgttcttgaaaaaacatcaccatcagatgatcatccatccatctcgtgcga 1471
QY 1432 accagaataggcaaatlaagatctcgtgggtccagtgatctgtgtgtaagagccttggattcg 1491
   |||||||
Db 1472 accagaataggcaaatlaagatctcgtgggtccagtgatctgtgtgtaagagccttggattcg 1531
QY 1492 aggcagctgagctcctcactgacagcagatgacgtcctgtgtactccagcctccttggga 1551
   |||||||
Db 1532 aggcagctgagctcctcactgacagcagatgacgtcctgtgtactccagcctccttggga 1591
QY 1552 aaccacagctctcctgtctccagtgaaagacttgatgagcagcaccatcgagggaagctgggt 1611
   |||||||
Db 1592 aaccacagctctcctgtctccagtgaaagacttgatgagcagcaccatcgagggaagctgggt 1651
QY 1612 cccagctgggaggtatggtgtgagctctatagaccatccctctctgcaatcaataaacac 1671
   |||||||
Db 1652 cccagctgggaggtatggtgtgagctctatagaccatccctctctgcaatcaataaacac 1711
QY 1672 ttgcctgtg 1680
   |||||||
Db 1712 ttgcctgtg 1720
   |||||||
```

```
RESULT 9
ID AAS72707 standard; cDNA; 2121 BP.
XX
AC AAS72707;
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #8511.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.
```

```
XX
OS Homo sapiens.
PN W0200175067-A2.
PD 11-OCT-2001.
PE 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT,
DR WPI; 2001-639362/73.
DR P-PSDB; ABG08520.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 8511; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging or sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 2121 BP; 515 A; 583 C; 561 G; 462 T; 0 other;
Query Match 63.3%; Score 1064; DB 23; Length 2121;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1284; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

```
QY 330 atccctgtgtcgtggcagcctgttgaaacgctcctgaaagcaccatcctgtgtgaagtc 389
   |||||||
Db 797 atccctgtgtcgtggcagcctgttgaaacgctcctgaaagcaccatcctgtgtgaagtc 856
QY 390 atcacagctaacatcctcagctcaggttgaaagccttcggcgaatgaccagagctgcta 449
   |||||||
Db 857 atcacagctaacatcctcagctcaggttgaaagccttcggcgaatgaccagagctgcta 916
QY 450 gtcaagatccccctgagacatgtgtgtgattcaacaacgcccctgtgcaagacatcgtg 509
   |||||||
Db 917 gtcaagatccccctgagacatgtgtgtgattcaacaacgcccctgtgcaagacatcgtg 976
QY 510 gagttcacatgacgacttgagcccaagcaccatccgcgatggaaccagtgcaagtggc 569
   |||||||
Db 977 gagttcacatgacgacttgagcccaagcaccatccgcgatggaaccagtgcaagtggc 1036
QY 570 cccacccgctgtcctcactgactgtgcccacagcaccatgggagcctgtgcacatcgaactg 629
   |||||||
```

Db 1037 ccacaccgcttggctccacgtgactgtgcccacagccatggagcctgcgcatccactg 1096
QY 630 ctgcataagctctccttcctgtgtgaagcgcttagcctaagcagtgataagactctccagt 689
Db 1097 ctgcataaagctctccttcctgtgtgaagcgcttagcctaagcagtgataagactctccagt 1156
QY 690 ccacccctgcgccaatctaagtgaanaaacgactgtgtcccgatgcaggtcttcctcaat 749
Db 1157 ccacccctgcgccaatctaagtgaanaaacgactgtgtcccgatgcaggtcttcctcaat 1216
QY 750 ggcattatgatcgaacctctcctgcagctgtgtgaaggtgtgccatttcctcagcatgtacgt 809
Db 1217 ggcattatgatcgaacctctcctgcagctgtgtgaaggtgtgccatttcctcagcatgtacgt 1276
QY 810 ctggaggttgacctctctgtatctctgcacaaaggtgtgaacacattcagcttcaactggag 869
Db 1277 ctggaggttgacctctctgtatctctgcacaaaggtgtgaacacattcagcttcaactggag 1336
QY 870 gccaaagtgttgactcaagggaaaggtgaccaaagtgttcaataactctgcagcttcc 929
Db 1337 gccaaagtgttgactcaagggaaaggtgaccaaagtgttcaataactctgcagcttcc 1396
QY 930 ctgcacaatgcccacccctggaacacatcccgcttcagccctcatcgtgtgaagcgtgtgtg 989
Db 1397 ctgcacaatgcccacccctggaacacatcccgcttcagccctcatcgtgtgaagcgtgtgtg 1456
QY 990 aaagctgcagctgtgctgtgtctctccagaagaattatgtgtctgttgcactctgtg 1049
Db 1457 aaagctgcagctgtgctgtgtctctccagaagaattatgtgtctgttgcactctgtg 1516
QY 1050 ctctctgaaagctgcacatcgctgaagctgaagctacatcgagctatcaataaagcgtgca 1109
Db 1517 ctctctgaaagctgcacatcgctgaagctgaagctacatcgagctatcaataaagcgtgca 1576
QY 1110 gataagctgtggaatctaccagaatcgtgtgaagatcttaactcagaacactccagag-ctttt 1168
Db 1577 gataagctgtggaatctaccagaatcgtgtgaagatcttaactcagaacactccagagctttt 1636
QY 1169 tatagaaccaaggcctgcgaaggctggccaaactgtatcgtcgtggaggtgttccctccag 1228
Db 1637 tatagaaccaaggcctgcgaaggctggccaaactgtatcgtcgtggaggtgttccctccag 1696
QY 1228 tgaagccctccgaccttgttcaacctgtgcatcgaagcagactcgaagcctcaattta 1288
Db 1697 tgaagccctccgaccttgttcaacctgtgcatcgaagcagactcgaagcctcaattta 1756
QY 1289 caccaaaagtgacaactatactcaacttgataaacaatcagctcgtatcgatccagct 1348
Db 1757 caccaaaagtgacaactatactcaacttgataaacaatcagctcgtatcgatccagct 1816
QY 1349 gatgaactctggattgtgtgtgttccaactgtatgttctgaaaacaatcatactcagagat 1408
Db 1817 gatgaactctggattgtgtgtgttccaactgtatgttctgaaaacaatcatactcagagat 1876
QY 1409 catccactcactcctgtgtgcgaacacagaatgtgcaaatlaagaatctcgtgggtccagtgac 1468
Db 1877 catccactcactcctgtgtgcgaacacagaatgtgcaaatlaagaatctcgtgggtccagtgac 1936
QY 1469 attgtgtgaagcctcgtggatctcgaagcagctgtatcctcactgacccaagaatgcctgtg 1528
Db 1937 attgtgtgaagcctcgtggatctcgaagcagctgtatcctcactgacccaagaatgcctgtg 1996
QY 1529 gcttactcagcctcctctgtgaaacccagctctcctgtctccagtgaaagctgtgag 1588
Db 1997 gcttactcagcctcctctgtgaaacccagctctcctgtctccagtgaaagctgtgag 2056
QY 1589 gcagccatcagggaaggtcgtgggtccca 1615
Db 2057 gcagccatcagggaaggtcgtgggtccca 2083

RESULT 10
ABAO1989

ID ABAO1989 standard; cDNA; 1035 BP.
AC ABAO1989;
XX
DT 12-FEB-2002 (first entry)
XX
DE Human NOV6b coding sequence.
XX
KW Human; NOV6b; metabolic disorder; neurodegenerative disorder;
KW immune disorder; haematopoietic disorder; developmental disease; cancer;
KW retinal disease; feeding disorder; vaccine; infection; gene therapy;
KW neurological disorder; psychotic disorder; G-protein coupled receptor;
KW cystostatic; antidiabetic; virucide; neuroprotective; nootropic;
KW analgesic; antidepressant; antimigraine; anticonvulsant; neuroleptic;
KW antispasmodic; antiatherosclerotic; antibacterial; fungicide;
KW osteopahtic; protozoacide; antifungal; hypertensive;
KW antifertility; vulnery; nephrotoxic; antileptic;
KW salivary gland protein; chromosome 20; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..79
FT /*tag= a
FT 79..1035
FT CDS /*tag= b
FT /*product= "NOV6b"
FT sig_peptide 79..150
FT /*tag= c
FT mat_peptide 151..1032
FT /*tag= d
XX
PM WO200174851-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US10039.
XX
PR 30-MAR-2000; 2000US-193205P.
PR 30-MAR-2000; 2000US-193339P.
PR 05-APR-2000; 2000US-195343P.
PR 06-APR-2000; 2000US-195005P.
PR 06-APR-2000; 2000US-195088P.
PR 10-APR-2000; 2000US-195792P.
PR 11-APR-2000; 2000US-196556P.
PR 13-APR-2000; 2000US-197081P.
PR 14-APR-2000; 2000US-197087P.
PR 14-APR-2000; 2000US-197525P.
PR 29-MAR-2001; 2001US-0823187.
XX
PA (CUBA-) CUBAGEN CORP.
XX
PI Majumder K, Spaderna SK, Taupier RU, Padigaru M, Burgess CE;
PI Shinkets RA, Spytek KA, Liu X, Patturajan M, Gusev VY;
XX
DR WPI: 2001-626379/72.
DR P-PSDB: AAM47220.
XX
PT New G protein-coupled receptor related polypeptides and polynucleotides
PT for diagnosis, prevention and treatment of metabolic,
PT neurodegenerative, retinal, immune, hematopoietic disorders, diabetes,
PT obesity and infections -
XX
PS Claim 8; Page 57; 194pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC human G-protein coupled receptors, designated NOV1, NOV2, NOV3, NOV4,
CC NOV5a, NOV5b, NOV6a, NOV6b, NOV7, NOV8a, NOV8b, NOV9 and NOV10. These can
CC be used in the treatment of NOVX related diseases, including cancer,
CC metabolic, neurodegenerative, immune, haematopoietic, developmental,
CC retinal, feeding, neurological and psychotic diseases and disorders and
CC infections. The present sequence is the NOV6b coding sequence, the gene

CC for which is found on chromosome 20. The NOV6b protein shares homology
CC with salivary gland proteins.
XX
SQ Sequence 1035 BP; 230 A; 333 C; 261 G; 211 T; 0 other;

Query Match 33.5%; Score 562; DB 22; Length 1035;
Best Local Similarity 99.9%; Pred. No. 4.1e-262;
Matches 662; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 102 ctgacctctgacacctggaagatgagccgctgagaccttaccctctctgtgttt 161
DB 57 ctgacctctgacacctggaagatgagccgctgagaccttaccctctctgtgttt 116
QY 162 gctgagcagcactctgtatccaagccacttcagctccactgtacttcatctcgtgcc 221
DB 117 gctgagcagcactctgtatccaagccacttcagctccactgtacttcatctcgtgcc 176
QY 222 aagaatcatcaagaagaactgacacagagctgaagagccacacgcccacagatcct 281
DB 177 aagaatcatcaagaagaactgacacagagctgaagagccacacgcccacagatcct 236
QY 282 gcaagagctgcgctgctcagtgatgcctgaggaagagcagcagagg-attccctgtgt 340
DB 237 gcaagagctgcgctgctcagtgatgcctgaggaagagcagcagaggatccctgtgt 296
QY 341 gggcagcctggtgaaacacctctctgaaagacatcatctgtgtgaagtcacacagctaa 400
DB 297 gggcagcctggtgaaacacctctctgaaagacatcatctgtgtgaagtcacacagctaa 356
QY 401 catctccagctgacagtgaaagcctcggccaatgacacagagctgtcagtcaagatccc 460
DB 357 catctccagctgacagtgaaagcctcggccaatgacacagagctgtcagtcaagatccc 416
QY 461 cctggaacatgtgtgtgattcaacacgcccctgtgtcaagacatctgtgagttccat 520
DB 417 cctggaacatgtgtgtgattcaacacgcccctgtgtcaagacatctgtgagttccat 476
QY 521 gacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 580
DB 477 gacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 536
QY 581 ggtctcagtgactgtgtccaccagcatalggagcctgagcagcagcagcagcagcagc 640
DB 537 ggtctcagtgactgtgtccaccagcatalggagcctgagcagcagcagcagcagcagc 596
QY 641 ctctctcctggtgaaagccttaagtaagcagtgatgaacctctcctagtcacatccctgg 700
DB 597 ctctctcctggtgaaagccttaagtaagcagtgatgaacctctcctagtcacatccctgg 656
QY 701 caatctagtgaaaaacagcagtggtcccgatcgagagcttcttcaatgagcatatgc 760
DB 657 caatctagtgaaaaacagcagtggtcccgatcgagagcttcttcaatgagcatatgc 716
QY 761 agacctctctcagcagtgatgaag 783
DB 717 agacctctctcagcagtgatgaag 739

RESULT 11
AAFI8254
ID AAF18254 standard; DNA; 1177 BP.
XX
AC AAF18254;

XX 14-MAR-2001 (first entry)

DE Lung cancer associated polynucleotide sequence SEQ ID 273.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnery;
KW gastrointestinal; nephrotoxic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

KW proliferative disorder; wound healing; infectious disease; ds.
XX
XX Homo sapiens.
OS
XX
XX WO20005180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM:
XX
XX WPI; 2000-587514/55.
XX P-PDB; AAB58378.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
XX antigens, useful for treatment, prevention, and diagnosis of disorders
XX such as lung cancer -
XX
XX Claim 1; Page 732; 1425pp; English.

CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
XX sequences.

SQ Sequence 1177 BP; 290 A; 361 C; 287 G; 239 T; 0 other;

Query Match 31.1%; Score 523; DB 21; Length 1177;
Best Local Similarity 99.8%; Pred. No. 3.2e-243;
Matches 643; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 53 ggaagcggcggagacacacagcagcagcagcagcagcagcagcagcagcagcagcagc 112
DB 2 ggaagcggcggagacacacagcagcagcagcagcagcagcagcagcagcagcagcagc 61
QY 113 caactggaagatgagccgagcagcagcagcagcagcagcagcagcagcagcagcagc 172
DB 62 caactggaagatgagccgagcagcagcagcagcagcagcagcagcagcagcagcagc 121
QY 173 cctgtatccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 232
DB 122 cctgtatccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 181
QY 233 aagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 292
DB 182 aagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 241
QY 293 cgtgtctcagtgccatgagcagcagcagcagcagcagcagcagcagcagcagcagc 351
DB 242 cgtgtctcagtgccatgagcagcagcagcagcagcagcagcagcagcagcagcagc 301
QY 352 tgacaacgcgtctgaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 411

|||||
Db 302 tgaacacgcgtccctgaagacacatcatctgctgtaaggtcatcaagctaacatctccagc 361
QY 412 tgaaggtgaagcctggtgccaatgaccaggaagctgtctatgaagatccccctggaatg 471
Db 362 tgaaggtgaagcctggtgccaatgaccaggaagctgtctatgaagatccccctggaatg 421
QY 472 tgcgtgatccaacagccctgtgtcaagaccatctgtgagttccatgaagacaggaag 531
Db 422 tgcgtgatccaacagccctgtgtcaagaccatctgtgagttccatgaagacaggaag 481
QY 532 cccaagccaacatccgcagctgagacacagctgcaagtgccaccgcgctgttctcaagt 591
Db 482 cccaagccaacatccgcagctgagacacagctgcaagtgccaccgcgctgttctcaagt 541
QY 592 actgtgcaaccagcagctgagagcctgtgcacatccactgtctgataagctctctctg 651
Db 542 actgtgcaaccagcagctgagagcctgtgcacatccactgtctgataagctctctctg 601
QY 652 tgaacgccttagcctaagcagatcatgaacctctagtgccatcc 695
Db 602 tgaacgccttagcctaagcagatcatgaacctctagtgccatcc 645

RESULT 12
AA297023
ID AA297023 standard; cDNA; 1175 BP.
AC AA297023;

DT 19-APR-2000 (first entry)
DE Human secreted protein gene 5 cDNA clone HBHMA23, SEQ ID NO:15.
XX

KW Human; secreted protein; cancer; tumour; developmental abnormality;
KW foetal deficiency; blood disorder; immune system disorder; inflammation;
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
KW therapy; ds.

XX Homo sapiens.
XX OS
XX PN MO9966041-A1.
XX
PD 23-DEC-1999.
XX
PF 15-JUN-1999; 99WO-US13418.
XX
PR 16-JUN-1998; 98US-0089507.
PR 16-JUN-1998; 98US-0089508.
PR 16-JUN-1998; 98US-0089509.
PR 16-JUN-1998; 98US-0089510.
PR 22-JUN-1998; 98US-0090112.
PR 22-JUN-1998; 98US-0090113.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA;
PI Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R;
PI Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;
XX
DR MPI: 2000-106100/09.
P-PSDB: AAY86219.

XX New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders -
XX
PS Claim 1: Page 318; 586pp: English.
XX
CC AA297019 to AA297137 represent 94 isolated human secreted protein genes.

CC AAY86215 to AAY86333 are the secreted proteins encoded by the 94 human
CC genes. This sequence represents a fragment of one of the human secreted
CC proteins. The genes and their corresponding secreted polypeptides are
CC useful for preventing, treating or ameliorating medical conditions,
CC e.g., by protein or gene therapy. Also pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new genes. Specific
CC uses are described for each of the 94 genes, based on which tissues they
CC are most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, tumours, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC disorders, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The sequences shown in AAY86334 to AAY86585 represent fragments of the
CC secreted proteins.
XX
SQ Sequence 1175 BP; 290 A; 359 C; 286 G; 239 T; 1 other;

Query Match 31.1%; Score 522; DB 21; Length 1175;
Best Local Similarity 99.8%; Pred. No. 9,9e-243;
Matches 642; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 54 gaagcgggcccgaagactccacgcgtgcccaggtcgtgacatccctgacccctctgac 113
Db 1 gaagcgggcccgaagactccacgcgtgcccaggtcgtgacatccctgacccctctgac 60
QY 114 acctgggaagatgcccgcgcgtggaaccttacccttctctgtgttctggaagccac 173
Db 61 acctgggaagatgcccgcgcgtggaaccttacccttctctgtgttctggaagccac 120
QY 174 ctgtatccaagccaccctcagttccatcagttcattctctctgcccacaagatcatca 233
Db 121 ctgtatccaagccaccctcagttccatcagttcattctctctgcccacaagatcatca 180
QY 234 agaaagagctgaacagagagcttgaagacacacacgcccacagatccctggaagctgac 293
Db 181 agaaagagctgaacagagagcttgaagacacacacgcccacagatccctggaagctgac 240
QY 294 gctgtcagtgccatcgaggaaagccagccgagag-atccctgtgctgggcagctggt 352
Db 241 gctgtcagtgccatcgaggaaagccagccgagagatccctgtgctgggcagctggt 300
QY 353 gaacacgcgtctgaagacacatctgtgctgaagttcatcaagctaacatctctcagct 412
Db 301 gaacacgcgtctgaagacacatctgtgctgaagttcatcaagctaacatctctcagct 360
QY 413 gcaagtgtaagcctctggccaatgaacagagagctgctgaataatcccccttgacatggt 472
Db 361 gcaagtgtaagcctctggccaatgaacagagagctgctgaataatcccccttgacatggt 420
QY 473 gcttgatccaacagccctctgttcaagacacatctgtgagttcacaatgaacagctgagc 532
Db 421 gcttgatccaacagccctctgttcaagacacatctgtgagttcacaatgaacagctgagc 480
QY 533 ccaagccaacatccgcagatgagacacagctgcaatgtgcccaccgcgctgttctcaagta 592
Db 481 ccaagccaacatccgcagatgagacacagctgcaatgtgcccaccgcgctgttctcaagta 540
QY 593 ctgtgcaaccagcagctgagagcctgtgacatccaactgtgcatataagctctctctcgtgt 652
Db 541 ctgtgcaaccagcagctgagagcctgtgacatccaactgtgcatataagctctctctcgtgt 600
QY 653 gaacgccttagcctaagcagatcatgaacctctcagtgccatcc 695
Db 601 gaacgccttagcctaagcagatcatgaacctctcagtgccatcc 643

RESULT 13
AA297114

ID AA297114 standard; cDNA; 1172 BP.
AC AA297114;
DT 19-APR-2000 (first entry)
XX
XX Human secreted protein gene 5 cDNA clone HBHMA23, SEQ ID NO:106.
PI
KW Human; secreted protein; cancer; tumour; developmental abnormality;
KW foetal deficiency; blood disorder; immune system disorder; inflammation;
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
KW therapy; ds.
XX
OS Homo sapiens.
PN W09966041-A1.
XX
PD 23-DEC-1999.
XX
PF 15-JUN-1999; 99WO-US13418.
XX
PR 16-JUN-1998; 98US-0089507.
PR 16-JUN-1998; 98US-0089508.
PR 16-JUN-1998; 98US-0089509.
PR 16-JUN-1998; 98US-0089510.
PR 22-JUN-1998; 98US-0090112.
PR 22-JUN-1998; 98US-0090113.
XX
PA (HDNA-) HUMAN GENOME SCI INC.
PI Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA,
PI Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Edner R;
PI Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;
DR WPT: 2000-106100/09.
XX P-PSDB: AAY86310.
XX
XX New isolated human genes and the secreted polypeptides they encode,
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders -
XX
XX Claim 1; Page 374; 586pp; English.

AA297019 to AA297137 represent 94 isolated human secreted protein genes.
AA86215 to AAY86333 are the secreted proteins encoded by the 94 human
genes. This sequence represents a fragment of one of the human secreted
proteins. The genes and their corresponding secreted polypeptides are
useful for preventing, treating or ameliorating medical conditions,
e.g., by protein or gene therapy. Also pathological conditions can be
diagnosed by determining the amount of the new polypeptides in a sample
or by determining the presence of mutations in the new genes. Specific
uses are described for each of the 94 genes, based on which tissues they
are most highly expressed in, and include developing products for the
diagnosis or treatment of cancer, tumours, developmental abnormalities
and foetal deficiencies, blood disorders, tumours, developmental
autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
disorders, digestive/endocrine disorders, infections and AIDS. The
polypeptides are also useful for identifying their binding partners.
The sequences shown in AAY86334 to AAY86585 represent fragments of the
secreted proteins.

Sequence 1172 BP; 287 A; 358 C; 201 G; 238 T; 8 other;

[illegible]

Query Match 24.1%; Score 405; DB 22; Length 963;
 Best Local Similarity 100.0%; Pred. No. 5.1e-186;
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 379 ggcgtgaaggtatcatcagatcaatctctcagctgcaggtgaagcctcggccaatgacc 438
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 263 ggcgtgaaggtatcatcagatcaatctctcagctgcaggtgaagcctcggccaatgacc 322
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 439 aggaagctctagtcagaatccccctggacatggtggtctggaattcaacacggccctgtgca 498
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 323 aggaagctctagtcagaatccccctggacatggtggtctggaattcaacacggccctgtgca 382
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 499 agaccatcgtggagttccacatgaagactgaagcccaagccacatccgcattggacacca 558
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 383 agaccatcgtggagttccacatgaagactgaagcccaagccacatccgcattggacacca 442
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 559 gtgcaagtggcccaagcccgctgtgctcctcaagtgaactgtgccacagccatlggaagctgc 618
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 443 gtgcaagtggcccaagcccgctgtgctcctcaagtgaactgtgccacagccatlggaagctgc 502
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 619 gcatccaactgtgtgataagctctcctcctgtgtgaagccttagtgaagcaggtcatga 678
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 503 gcatccaactgtgtgataagctctcctcctgtgtgaagccttagtgaagcaggtcatga 562
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 679 acctcctagtgccatccctcctgccaatctagtgaanaaaccaagctgtgtcccggtgacgag 738
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 563 acctcctagtgccatccctcctgccaatctagtgaanaaaccaagctgtgtcccggtgacgag 622
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 739 ctctcctcaatggatgtatgcagacctcctcctgcaagctgtgtgaag 783
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 623 ctctcctcaatggatgtatgcagacctcctcctgcaagctgtgtgaag 667
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    
```

Search completed: September 16, 2002, 06:31:11
 Job time: 7619 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:27:07 ; Search time 90.91 Seconds

(without alignments)
4539.262 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 ggtgtcagagataaagtt.....tcaataacactgtcctgtg 1680

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 38353 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PC10S.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 21 | 1.2 | 7301 | 4 | US-09-816-088-3 |
| 2 | 19 | 1.1 | 3471 | 2 | US-08-715-568A-2 |
| 3 | 18 | 1.1 | 1406 | 1 | US-08-182-060A-6 |
| 4 | 18 | 1.1 | 1406 | 1 | US-08-712-702A-6 |
| 5 | 18 | 1.1 | 2068 | 2 | US-08-466-589-1 |
| 6 | 18 | 1.1 | 2068 | 2 | US-08-700-636-1 |
| 7 | 18 | 1.1 | 2068 | 3 | US-08-467-574-1 |
| 8 | 18 | 1.1 | 2068 | 4 | US-09-217-345-1 |
| 9 | 18 | 1.1 | 2277 | 1 | US-08-496-855A-1 |
| 10 | 18 | 1.1 | 3033 | 1 | US-08-453-695A-111 |
| 11 | 18 | 1.1 | 3033 | 1 | US-08-268-161A-111 |
| 12 | 18 | 1.1 | 3033 | 2 | US-08-453-702A-111 |
| 13 | 18 | 1.1 | 3033 | 4 | US-09-099-639-111 |
| 14 | 18 | 1.1 | 3033 | 5 | PCT-US95-08071-111 |
| 15 | 18 | 1.1 | 3412 | 4 | US-08-903-325-3 |
| 16 | 18 | 1.1 | 11459 | 4 | US-09-454-721A-3 |
| 17 | 17 | 1.0 | 864 | 1 | US-08-396-650-2 |
| 18 | 17 | 1.0 | 864 | 1 | US-08-768-626-2 |
| 19 | 17 | 1.0 | 864 | 2 | US-08-875-062-2 |
| 20 | 17 | 1.0 | 921 | 1 | US-08-396-650-3 |
| 21 | 17 | 1.0 | 921 | 1 | US-08-396-650-4 |
| 22 | 17 | 1.0 | 921 | 1 | US-08-768-626-3 |
| 23 | 17 | 1.0 | 921 | 1 | US-08-768-626-4 |
| 24 | 17 | 1.0 | 1244 | 5 | PCT-US96-02331-12 |
| 25 | 17 | 1.0 | 1355 | 3 | US-08-415-655-14 |
| 26 | 17 | 1.0 | 1575 | 1 | US-08-090-523-7 |
| 27 | 17 | 1.0 | 1575 | 1 | US-08-398-627-7 |

| | | | | | | |
|----|----|-----|---------|---|-------------------|--------------------|
| 28 | 17 | 1.0 | 1575 | 1 | US-08-406-858-7 | Sequence 7, Appli |
| 29 | 17 | 1.0 | 1575 | 5 | PCT-US91-04036-7 | Sequence 7, Appli |
| 30 | 17 | 1.0 | 1575 | 5 | PCT-US94-05275-7 | Sequence 7, Appli |
| 31 | 17 | 1.0 | 2711 | 4 | US-08-891-298-2 | Sequence 2, Appli |
| 32 | 17 | 1.0 | 4255 | 5 | PCT-US96-02331-14 | Sequence 14, Appli |
| 33 | 17 | 1.0 | 8791 | 5 | PCT-US96-01735-5 | Sequence 5, Appli |
| 34 | 17 | 1.0 | 4411529 | 4 | US-09-103-840A-1 | Sequence 1, Appli |
| 35 | 17 | 1.0 | 27 | 5 | PCT-US94-01263-4 | Sequence 4, Appli |
| 36 | 16 | 1.0 | 50 | 3 | US-08-825-852-22 | Sequence 22, Appli |
| 37 | 16 | 1.0 | 50 | 4 | US-09-052-888-22 | Sequence 22, Appli |
| 38 | 16 | 1.0 | 58 | 3 | US-08-825-852-23 | Sequence 23, Appli |
| 39 | 16 | 1.0 | 58 | 4 | US-09-052-888-23 | Sequence 23, Appli |
| 40 | 16 | 1.0 | 81 | 4 | US-08-556-978B-27 | Sequence 27, Appli |
| 41 | 16 | 1.0 | 81 | 4 | US-08-556-978B-28 | Sequence 28, Appli |
| 42 | 16 | 1.0 | 99 | 4 | US-09-199-149-24 | Sequence 24, Appli |
| 43 | 16 | 1.0 | 192 | 2 | US-08-442-809A-49 | Sequence 49, Appli |
| 44 | 16 | 1.0 | 285 | 1 | US-08-236-520-4 | Sequence 4, Appli |
| 45 | 16 | 1.0 | 285 | 5 | PCT-US95-05262-4 | Sequence 4, Appli |

ALIGNMENTS

```

RESULT 1
US-09-816-088-3
; Sequence 3, Application US/09816088
; Patent No. 6326180
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-hui et al
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME, NUCLEIC ACID
; FILE REFERENCE: CLO01181
; CURRENT APPLICATION NUMBER: US/09/816,088
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7301
; TYPE: DNA
; ORGANISM: Human
US-09-816-088-3

Query Match 1.2% Score 21; DB 4; Length 7301;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 785 gccacatccctcagcatgca 805
Db 3365 gccacatccctcagcatgca 3385

RESULT 2
US-08-715-568A-2
; Sequence 2, Application US/08715568A
; Patent No. 5856463
; GENERAL INFORMATION:
; APPLICANT: Prydz, Hans Peter Blankenborg
; APPLICANT: Brede, Gaute
; TITLE OF INVENTION: PSKH-1 Ribozymes and Uses in Disease
; TITLE OF INVENTION: Treatment
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
; STREET: 600 South Avenue West
; CITY: Westfield
; STATE: NJ
; COUNTRY: USA
; ZIP: 07090-1497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/715,568A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Foley, Shawn P.
REFERENCE/DOCKET NUMBER: FORSK 3.0-002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-654-5000
TELEFAX: 908-654-7866
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3471 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-715-568A-2

Query Match
Best Local Similarity 1.1%; Score 19; DB 2; Length 3471;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 134 cgtggaccttcacctct 152
|||||
DB 2070 CGTGGACCTTCACCTTCT 2088

RESULT 3
US-08-182-060A-6
Sequence 6, Application US/08182060A
Patent No. 5648210
GENERAL INFORMATION:
APPLICANT: John W. Pierce
APPLICANT: Phillip S. Kerr
APPLICANT: Mary F. Becker-Manley
APPLICANT: Richard W. Pearlstein
APPLICANT: Bruce J. Schweiger
TITLE OF INVENTION: Nucleotide Sequences
TITLE OF INVENTION: of Galactinol
TITLE OF INVENTION: Synthase from
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours
ADDRESSEE: and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,060A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 92/06057
FILING DATE: 24 JULY 1992
APPLICATION NUMBER: 07/735,066
FILING DATE: 24 JULY 1991
ATTORNEY/AGENT INFORMATION:
NAME: LINDA AXAMETHY FLOYD
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1032-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302)992-4929
TELEFAX: (302)892-7949

TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1406 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Glycine max
STRAIN: Cultivar Wye
IMMEDIATE SOURCE:
LIBRARY: cDNA to mRNA
CLONE: ps21
PUBLICATION INFORMATION: Unpublished
US-08-182-060A-6

Query Match
Best Local Similarity 1.1%; Score 18; DB 1; Length 1406;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1394 catcatcagagatcat 1411
|||||
DB 123 CATCATCCTGACATCAT 140

RESULT 4
US-08-712-702A-6
Sequence 6, Application US/08712702A
Patent No. 5773699
GENERAL INFORMATION:
APPLICANT: John W. Pierce
APPLICANT: Phillip S. Kerr
APPLICANT: Mary F. Becker-Manley
APPLICANT: Richard W. Pearlstein
APPLICANT: Bruce J. Schweiger
TITLE OF INVENTION: Nucleotide Sequences
TITLE OF INVENTION: of Galactinol Synthase from Zucchini and Soybean
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours
ADDRESSEE: and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/712,702A
FILING DATE: 12-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/182,060
FILING DATE: 21-JAN-1994
APPLICATION NUMBER: US 92/06057
FILING DATE: 24 JULY 1992
APPLICATION NUMBER: 07/735,066
FILING DATE: 24 JULY 1991
ATTORNEY/AGENT INFORMATION:
NAME: LINDA AXAMETHY FLOYD
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1032-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302)992-4929
TELEFAX: (302)892-7949
TELEX:
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 1406 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Glycine max
SPRAIN: Cultivar Wye
IMMEDIATE SOURCE:
LIBRARY: cDNA to mRNA
CLONE: PS21
PUBLICATION INFORMATION: Unpublished
US-08-712-702A-6

Query Match 1.1%; Score 18; DB 1; Length 1406;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1394 tcctcactgagatcat 1411
DB 123 CATCATCTGAGATCAT 140

RESULT 5
US-08-466-589-1
Sequence 1, Application US/08466589
Patent No. 5837489
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 12
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,589
FILING DATE: June 5, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: March 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9950
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2068 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 166..1752

US-08-466-589-1

Query Match 1.1%; Score 18; DB 2; Length 2068;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tcctcactgagatcatcc 1413
DB 1100 TCATCATCTGAGATCATCC 1117

RESULT 6
US-08-700-636-1
Sequence 1, Application US/08700636
Patent No. 5910582
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 12
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/700,636
FILING DATE: 16-JUL-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: 08-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9368
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2068 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 166..1752
US-08-700-636-1

Query Match 1.1%; Score 18; DB 2; Length 2068;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tcctcactgagatcatcc 1413
DB 1100 TCATCATCTGAGATCATCC 1117

RESULT 7
US-08-467-574-1

; Sequence 1, Application US/08467574
; Patent No. 6022704
; GENERAL INFORMATION:
; APPLICANT: Elliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,574
; FILING DATE: June 5, 1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/028,031
; FILING DATE: March 8, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9949
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 166..1752
; US-08-467-574-1

Query Match 1.1%; Score 18; DB 3; Length 2068;
Best Local Similarity 100.0%; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tcatactgagatcacc 1413
DB 1100 TCATCCTGAGATCATCC 1117

RESULT 8
; Sequence 1, Application US/09217345
; Patent No. 6303753
; GENERAL INFORMATION:
; APPLICANT: Elliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McCauliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: LA Jolla

; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/217,345
; FILING DATE: 21-DEC-98
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/467,574
; FILING DATE: 05-JUN-95
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/466,589,
; FILING DATE: 05-JUN-95
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/028,031
; FILING DATE: 08-MAR-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24735-9949B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-450-8400
; TELEFAX: 619-587-5360
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 166..1752
; US-09-217-345-1

Query Match 1.1%; Score 18; DB 4; Length 2068;
Best Local Similarity 100.0%; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tcatactgagatcacc 1413
DB 1100 TCATCCTGAGATCATCC 1117

RESULT 9
; Sequence 1, Application US/08496855A
; Patent No. 5801232
; GENERAL INFORMATION:
; APPLICANT: Elliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,855A
FILING DATE: 20-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,503
FILING DATE: 08-NOV-1993
APPLICATION NUMBER: US 08/028,031
FILING DATE: 08-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9369B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 166..1755
US-08-496-855A-1

Query Match 1.1%; Score 18; DB 1; Length 2277;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tcactcagagatcc 1413
DB 1103 tcactcagagatcc 1120

RESULT 10
US-08-453-695A-111
Sequence 111, Application US/08453695A
Patent No. 5708143
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
ADDRESS: Borun
STREET: 233 South Wacker, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,695A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: No. 5708143and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32658
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:

LENGTH: 3033 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 138..2528
US-08-453-695A-111

Query Match 1.1%; Score 18; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1051 ttcctgagagtgccatc 1068
DB 562 TTCCTGAGAGTGCCATC 579

RESULT 11
US-08-268-161A-111
Sequence 111, Application US/08268161A
Patent No. 5798224
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
ADDRESS: Borun
STREET: 233 South Wacker, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/268,161A
FILING DATE: June 27, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Young J. Sun
REGISTRATION NUMBER: P-41,337
REFERENCE/DOCKET NUMBER: 27866/32149
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 3033 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 138..2528
US-08-268-161A-111

QY 1051 ttcctgagagtgccatc 1068
DB 562 TTCCTGAGAGTGCCATC 579

RESULT 12
US-08-453-702A-111
; Sequence 111, Application US/08453702A
; Patent No. 5891706
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,702A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5891706and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32657
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 138..2528
US-08-453-702A-111

Query Match 1.1%; Score 18; DB 2; Length 3033;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1051 ttctgtagatgcccac 1068
|||||

Db 562 TTCCTGAGAGTCCCATC 579

RESULT 13
US-09-099-639-111
; Sequence 111, Application US/09099639
; Patent No. 6262237
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/099,639
; FILING DATE: 18 JUN 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/263,161
; FILING DATE: 27 JUN 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Greta E. No. 6262237and
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/34703
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 138..2528
US-09-099-639-111

Query Match 1.1%; Score 18; DB 4; Length 3033;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1051 ttctgtagatgcccac 1068
|||||

Db 562 TTCCTGAGAGTCCCATC 579

RESULT 14
PCT-US95-08071-111
; Sequence 111, Application PC/TUS9508071
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08071
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12588
; FILING DATE: 23 DEC 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/998,003
; FILING DATE: 29 DEC 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.

```

;
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32149
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
;
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 138..2528
; PCT-US95-08071-111

Query Match 1.1%; Score 18; DB 5; Length 3033;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1051 ttcttgagagtgcccatc 1068
Db 562 TTCCTGAGAGTGCCCATC 579

RESULT 15
US-08-903-325-3
; Sequence 3, Application US/08903325
; Patent No. 6262338
; GENERAL INFORMATION:
; APPLICANT: Peter Schreier; Thomas Herget; and
; TITLE OF INVENTION: RESISTANCE GENES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG HORN KRAMER & WOODS
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; MEDIUM TYPE: storage
; COMPUTER: NEC Powermate SX/20
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/903,325
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/383,747
; FILING DATE: 27-JAN-1995
; APPLICATION NUMBER: DE 40 31 758.7
; FILING DATE: 06-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/766990
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235106
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kuit G. Biscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 8297.2-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:

```

```

;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3412 nucleotides
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Peanut (arachis hypogaea)
; US-08-903-325-3

Query Match 1.1%; Score 18; DB 4; Length 3412;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1475 gaagcccttgagattcga 1492
Db 2580 GAAGCCCTTGAGATTGCA 2597

```

Search completed: September 16, 2002, 06:24:15
Job time: 10628 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 02:58:27 : Search time 3017.73 Seconds
(without alignments)
7513.888 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 ggtgtgcagataaagtt.....tcataaacactgtcctgtg 1680

Scoring table:

Gapop 60.0, Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size: 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 662 | 39.4 | 783 | 10 | BI770190 603053221 |
| 2 | 636 | 37.9 | 807 | 10 | BI823956 603039139 |
| 3 | 599 | 35.7 | 852 | 10 | BI818205 603032679 |
| 4 | 588 | 35.0 | 784 | 10 | BI821790 603035865 |
| 5 | 551 | 33.8 | 955 | 10 | BI821769 603035838 |
| 6 | 490 | 29.2 | 524 | 10 | BE529820 602558904 |
| 7 | 485 | 28.9 | 658 | 9 | BE061388 QVO-BT002 |
| 8 | 451 | 26.8 | 567 | 10 | BE088537 RCL-HT088 |
| 9 | 446 | 26.5 | 945 | 10 | BI772722 603053235 |
| 10 | 428 | 25.5 | 549 | 10 | BE482718 602502690 |
| 11 | 418 | 24.9 | 602 | 10 | BE530024 602558870 |
| 12 | 400 | 23.8 | 660 | 10 | BE548906 602575739 |
| 13 | 396 | 23.6 | 524 | 10 | BE933204 RCL-HT088 |
| 14 | 392 | 23.3 | 539 | 10 | BI824243 603040625 |
| 15 | 390 | 23.2 | 421 | 10 | BI824243 603040625 |
| 16 | 386 | 23.0 | 465 | 10 | BI824243 603040625 |
| 17 | 384 | 22.9 | 482 | 10 | BE095122 PM3-UT005 |

| | | | | | |
|----|-----|------|------|----|---------------------|
| 18 | 377 | 22.4 | 1026 | 10 | BI821994 603040023 |
| 19 | 374 | 22.3 | 469 | 10 | BE743981 RCO-BT081 |
| 20 | 371 | 22.1 | 453 | 9 | AI924632 w57e12.x |
| 21 | 370 | 22.0 | 957 | 10 | BI822130 603039831 |
| 22 | 348 | 20.7 | 640 | 10 | BE482968 602502981 |
| 23 | 348 | 20.7 | 642 | 10 | BE672220 602150403 |
| 24 | 348 | 20.7 | 785 | 10 | BE530180 602558678 |
| 25 | 345 | 20.5 | 433 | 10 | BE826403 CM4-HN002 |
| 26 | 342 | 20.4 | 525 | 9 | AL602621 DKFZP686K |
| 27 | 336 | 20.0 | 438 | 10 | AM167650 xm47a11.x |
| 28 | 331 | 19.7 | 370 | 10 | BE926678 QV2-BT063 |
| 29 | 323 | 19.2 | 390 | 10 | BE829540 RCO-ET007 |
| 30 | 322 | 19.2 | 365 | 9 | AM381927 QV1-HT031 |
| 31 | 320 | 19.0 | 556 | 10 | BE933205 RCL-HT088 |
| 32 | 320 | 19.0 | 912 | 10 | BE484717 602505896 |
| 33 | 318 | 18.9 | 630 | 10 | BE482764 602502549 |
| 34 | 312 | 18.6 | 363 | 9 | AM167610 xm49e05.x |
| 35 | 305 | 18.2 | 379 | 10 | BE088529 RCL-HT088 |
| 36 | 303 | 18.0 | 429 | 9 | AM449208 UI-H-B13- |
| 37 | 296 | 17.6 | 482 | 10 | BE926857 QV1-BT067 |
| 38 | 290 | 17.3 | 343 | 9 | BE073529 RC5-BT055 |
| 39 | 282 | 16.8 | 353 | 10 | BE089104 CM0-HT091 |
| 40 | 277 | 16.5 | 743 | 10 | BE548174 602575330 |
| 41 | 273 | 16.2 | 836 | 10 | BI820886 603033924 |
| 42 | 264 | 15.7 | 397 | 10 | BE693858 RC5-BT055 |
| 43 | 257 | 15.3 | 428 | 10 | BE9332793 CM2-HT087 |
| 44 | 246 | 14.6 | 318 | 9 | AM996310 QV3-BN004 |
| 45 | 235 | 14.0 | 287 | 10 | BE804024 CM2-C1013 |

ALIGNMENTS

RESULT 1
BI770190 783 bp mRNA linear EST 25-SEP-2001
LOCUS 603053221F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5202893 5',
DEFINITION mRNA sequence.
ACCESSION BI770190
VERSION BI770190.1 GI:15761768
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 783)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.fremail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL1508 row: 1 column: 06
High quality sequence stop: 782.
Location/Qualifiers
1. 783
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5202893"
/clone="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size

FEATURES

source

range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH-MGC Library."

BASE COUNT 167 a 261 c 199 g 156 t
ORIGIN

Query Match 39.4%; Score 662; DB 10; Length 783;
Best Local Similarity 99.9%; Pred. No. 0;

Matches 782; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

Oy 53 ggaagagccgagagcctccagcagctgcccagctgcaatccctgacattgctgcccctga 112
Db 1 ggaagccgcccagagcctccagcagctgcccagctgcaatccctgacattgctgcccctga 60
Oy 113 caccggaagatgagccgcccgtgagaccttccaccttctgtgtgtgtgtgtgtgtgtgtgt 172
Db 61 CACCTGGGAATATGGCCGGCCCGTGGACCTTCACCTCTCTGTGTGTGTGTGTGTGTGTGTGT 120
Oy 173 ccttgatcaagcaccctcagctccacatgctcctcctcctcctcctcctcctcctcctc 232
Db 121 CCTGATCCAAAGCACCCTCAGTCCCACTGCACTTCTCATCTCTGGCCCAAAATCATCA 180
Oy 233 aagaaaagctgacagagagctgaaagacacacacacacacacacacacacacacacacac 292
Db 181 AAGAAAGCTTACACAGAGAGCTGAAGACCAACAGCCACACAGCATCTGCAGCAGCTGC 240
Oy 293 cgtgctcagctgcatgctgaggaagacacacacacacacacacacacacacacacacac 351
Db 241 CGCTGCTCAGTGCATCGGGGAAAAGCCAGCGGAGGATCCCTGTCTGGGACGCTGG 300
Oy 352 tgaacacccctcctgaaagacacacacacacacacacacacacacacacacacacacac 411
Db 301 TGAACACCCGTCCTGAACACATCATCTGGCTGAAGCTCATCATCAAGCTAACATCTCCAGC 360
Oy 412 tgcagtggaagccctgagcaatgacagagagctgctgacagatgacacacacacacacac 471
Db 361 TGCAGGTGAAGCCCTCGGCCAATGACGAGAGCTGCTGATGATCAAGATCCCTCTGGACATGG 420
Oy 472 tgcgtgattcaacagccctcctgctcaagacacacacacacacacacacacacacacac 531
Db 421 TGGCTGATTCACACACGCCCTCTGCTCAAGACCATCTGAGTTCCACATGACCATGAGG 480
Oy 532 cccaagccacacatccgatgagacacacacacacacacacacacacacacacacacacacac 591
Db 481 CCCAAGCCACCATCCGATGAGACACCAAGTGAAGTGGCCCAACCCGCTGCTCCTGAGTG 540
Oy 592 actgtgacacacagcagctgagagcctgcacacacacacacacacacacacacacacacac 651
Db 541 ACTGTGACACACAGCATGGAGGCTCGCATTCACATCGCTGCTAAGCTCTCTCTCTGCG 600
Oy 652 tgaagccttaagtaagcagatcatgaacctcctagctgacacacacacacacacacacac 711
Db 601 TGAAGCCTTAAGTAAGCAGATCATGAACCTCTAGTGCATCCCTGACCAATCTAGTGA 660
Oy 712 aaacacagctgtgtccctgagatgagagcttctcctcaatgcatgtatgtgaagacacac 771
Db 661 AAACACAGCTGTCTCCCGTATGAGGCTTCTCTCAATGACATGATGAGACCTCTCTGCG 720
Oy 772 agctgtgaaagtgtccacattccctcagacattgacagctgtgagctgtgacattgtatc 831
Db 721 AGCTGTGAAAGTGCCCATTTCCCTCAGCATTGACCGCTGTGAGTTTGACCTTCTGTATTC 780
Oy 832 ctg 834
Db 781 CTG 783

```

RESULT 2
BI823956 807 bp mRNA linear EST 04-OCT-2001
LOCUS BI823956
DEFINITION 603039139P1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5180336 5',

ACCESSION mRNA sequence.
BI823956
VERSION BI823956.1 GI:15935506
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 807)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1149 row: m column: 09
High quality sequence stop: 793.

FEATURES
source Location/Qualifiers

1..807
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5180336"
/clone_11b="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site: 1; NotI; Site: 2; EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH-MGC Library."

BASE COUNT 175 a 269 c 202 g 161 t
ORIGIN

Query Match 37.9%; Score 636; DB 10; Length 807;
Best Local Similarity 99.9%; Pred. No. 6,76-314;

Matches 756; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

Oy 62 cgaagacccagcgtgcccagctgcatcctcctgacattgctgccccttgacacctgga 121
Db 7 CGAGGACTCCACCGCTGCCAGGTCTGGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 66
Oy 122 agatgcccgcgcgcgtgagacattcacccttctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 181
Db 67 AGATGGCCGGCCCGTGGACCTTCACCTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 126
Oy 182 aagccacccctcagctccacatgcatcctcctcctcctcctcctcctcctcctcctcctcct 241
Db 127 AAGCCACCCCTAGTCCCATCGATGCTTCATCTCTGCGCCCAAAAGTATCAAAAGAAAAGC 186
Oy 242 tgaacagagctgaaagacacacacacacacacacacacacacacacacacacacacacac 301
Db 187 TGACACAGAGCTGTAAGGACCAACAGCCACACACATCTCTGACAGAGTGCCGCTGTGCTCA 246
Oy 302 gtgcacatgcgggaaagacacacacacacacacacacacacacacacacacacacacacac 360
Db 247 GTGCCATCGGGGAAAGCCAGGACGATCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 306
Oy 361 tcttgaagcacatcatctgtgtgaaagtcacacacacacacacacacacacacacacacacac 420
Db 307 TCCTGAAGCACATCATCTGGCTGAAGGTCAATCAAGCTATACATCTCTCAGCTGAGGTGA 366
Oy 421 agccctcggcacaatgaccagagagctgtagtaagaatccccccttgacatggtgtgtgat 480

```

DB 367 ACCCTCGGCAATGACCGAGAGCTCTAGTCAGATCCCTGGACATGCTGGCTGAT 426
QY 481 tcaaacgcccctgttcaagaccatcgttgagttcccaatgacgactgagcccaagcca 540
DB 427 TCAGACGGCCCTGTGGCAAGACCATCTGTGAGTTCACATGACGACTGAGGCCCAAGCCA 486
QY 541 ccacccgacatggaacccaatgcaagtggcccaaccgctgtgtcccaatgactgtgcca 600
DB 487 CCATCCGGCATGGAACACATGCAAGTGGCCCAACCCGCTGTCTCTCAATGACTGTGCCA 546
QY 601 ccagccatgaggagcctgagccatcccaactgctgataaagctctcctcctgctggaacgct 660
DB 547 CCAGCATGAGGAGCTGCGCATTCACATCTGTCATGAGCTTCCTCCGAGGAGACGCT 606
QY 661 ttgctcaagagatcatgaacccctcctcctagtgccatccctgcccacatagtgtgaaaacacagc 720
DB 607 TTAGCTAAGAGGTTCATGAACCTCTCTAGTCCATCCCTGCCAATCTAGTGAAGAAACACAC 666
QY 721 ttgtcccgtagtcagagctcctcctcaatgcatgtatgacagcctcctcagctgtgta 780
DB 667 TGTGTCCCTGATCGAGGCTTCTTCAATGCGCATGATGACACCTCTCTGCAAGCTGTGA 726
QY 781 aggtgcccatctccctcagcatgacgctgaggt 817
DB 727 AGGTGCCATTTCCCTCAGCATTTGACCGCTGTGAGTT 763

RESULT 3

B1818205 852 bp mRNA linear EST 04-OCT-2001
LOCUS 603032679F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5173842 5',
DEFINITION mRNA sequence.

B1818205
ACCESSION B1818205.1 GI:15928527

KEYWORDS
SOURCE EST.

ORGANISM human.

Human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 852)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: c9apbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/MLN at:

http://image.llnl.gov

Plate: LLM11432 row: n column: 19

High quality sequence stop: 793.

FEATURES

source

1..852

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5173842"

/clone_11b="NIH_MGC_115"

/lab_host="DH10B"

/note="Organ: Pooled brain, lung, testis; Vector:

PCMV-SPORE; Site_1: NOTI; Site_2: EcoRV (destroyed); RNA

source anonymous pool of 6 male brains, age range 23-27; 1

male lung, age 27; and 1 male testis, age 69. Library is

oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.8 kb,

insert size range 1-3 kb. Library is normalized and

enriched for full-length clones and was constructed by C.

Gruber (Invitrogen). Research Genetics tracking code

021. Note: this is a NIH-MGC Library."

BASE COUNT 185 a 282 c 219 g 166 t

ORIGIN

Query Match 35.7%; Score 599; DB 10; Length 852;

Best Local Similarity 99.9%; Pred. No. 8,4e-295;

Matches 719; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 55 agcgggacgaagaccacccacgctgcccaggttcctgcatctcctgactgtgcctcctgaca 114
DB 1 AGCGGGCCGAGACATCCACAGCTGCCCCAGGTCTGGCATCTGCACATTCGCTGCCCCTGACA 60
QY 115 cctgggaagatgagccgagccgctgagccttcacacctctcctgtgtgttctgtgagcagcc 174
DB 61 CTTGGGAAGATGAGCCGGCCGGCCCGGAGACCTTCACACCTTCCTCTGTGAGTTGCTGGCAGCAGCAC 120
QY 175 ttgatccaaagcaccctcagctcagctcagctcagctcagctcagctcagctcagctcagct 234
DB 121 TTGATCCAAAGCCACCTTCAGTCCCACTGCAGTTCTCATCTCTGCGCCCAAAAGTCAATCAAA 180
QY 235 gaaaagctgacacagagctgaagac 294
DB 181 GAAAGCTGACACAGAGCTGAAGGACCAACAGCACCACGATCTTGCACAGCAGCTGCGG 240
QY 295 ctgtcagtgccatgacgaggaagacagcagcagcagcagcagcagcagcagcagcagcagcagc 353
DB 241 CTGCTCAGTGCATGCGGGAAGGCAAGCCGAGGAGCATCTGTGCTGGCAGCAGCTGCTG 300
QY 354 aacacccgtctgaagac 413
DB 301 AACACCGTCTTGAAGACATCATCTGGCTGAAGGTCATACAGCTAACATCTCTCCAGCTG 360
QY 414 caggtgaagcctcagcacaatgacagcagcagcagcagcagcagcagcagcagcagcagcagc 473
DB 361 CAGGTGAAGCCTCGGCCAATGACAGAGCTGCTAGTAAAGTCCCTTGACATGAGTG 420
QY 474 gcttgatltcaaacagccctctgttcaagacacatctgaggtltcacaatgacagcagcagc 533
DB 421 GCTGATTCAGACAGCGCCCTGTGGTCAAGACCATCTGGAATTCACATGACAGCATGAGGCC 480
QY 534 caagcaccatctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 593
DB 481 CAAGCCACCATCTCGCATGAGACACAGTGGACAGTGGCCCGCTGTGCTCTCATGTGAC 540
QY 594 ttgtccaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653
DB 541 TGTGCCACACGACCATGAGGAGCTGCGCATTCACATCTCTCTCTCTCTCTCTCTCTCTCTCT 600
QY 654 aacgccttagcctaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 713
DB 601 AACGCTTGTAGTAAGAGGTCATGAACCTCTAGTGCATCTCTCTCTCTCTCTCTCTCTCTCTCT 660
QY 714 aacgcagctgtgtccggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 773
DB 661 AACCAAGCTGTGTCCGTGATGAGGCTTCTCTCAATGAGCATGTATGCAAGACCTCTCTCTCTCT 720

RESULT 4

B1821790 784 bp mRNA linear EST 04-OCT-2001
LOCUS 603035865F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5176624 5',
DEFINITION mRNA sequence.

B1821790
ACCESSION B1821790.1 GI:15933340

KEYWORDS EST.

SOURCE human.

Human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 784)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

COMMENT

Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLML1440 row: b column: 17
High quality sequence stop: 763.
Location/Qualifiers

| BASE COUNT | ORIGIN |
|------------|--------|
| 168 a | 263 c |
| | 200 g |
| | 153 t |

| | | | | |
|---------------------------|--------|---------------------|-----------|-------------|
| Query Match | 35.0% | Score 588; | DB 10; | Length 784; |
| Best Local Similarity | 99.7%; | Pred. No. 3.5e-289; | | |
| Matches 758; Conservative | 0; | Mismatches 1; | Indels 1; | Gaps 1 |

55 agcggcgagagactccagcgctgccaggtctgacatcctgcacattgtgcctctgaca 114

115 cctggaagatgycgcgcctgacacttaccctctctgtgtgttctgtcgaacacac 174
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520

D61 CCTGGGAGATGGCCGGCCCGGTGACCTTCACCCCTCTGTGTGTTGCTGGCAGCCACC 120

173 ccgacaccagccaccctcagctccacacgagttctcatctcggcccaaaagtcataaa 234
|||||
121 ttgatccaaagccaccctcagctccacacgagttctcatctcggcccaaaagtcataaa 180
|||||

235 gaaagctgtacacagagagctggaagaccacaaacgcccacagcatcctgcagcagctgcg 294

295 ctgtcagtcgcatgcggaagaagccagccggagg-atccctgtgtcgtggcagcctgtg 353

[illegible]

Ddb 301 AACACCGTCTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCTCCAGCTG 360

[illegible]

4/4 gctggtatcacaacgcccctgttcaagaccatcgttgagtccatcatgtgacgtgaguc 533

534 caagccaccatccgatgacaccaagtgcgaatggccccaccgcctgctcctaagtac 593

401 CAGGCCACCAATCCGCAATGGACATCCAGTGGCAAGTGGCCCCACCCGCTGGTCTCTAGTGAC 540

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040

| | | | |
|----|-----|---|-----|
| Db | 541 | TTGGCCACGAGCATGSGAGSCTGGGCATCCAACTCTGCTAATAGCTCTCTCCTCGGTG | 600 |
| QY | 654 | aacgccttagctaagcaggtatgaacactctcctagtagcatalccctgccaatctagtga | 713 |
| Db | 601 | AAAGCCTTAGCTAAGCAGGTCATGAACCTCTAGTGCATCCCTGCCAATCTAGTGAA | 660 |
| QY | 714 | aacgccttagtaccgttagtcagagcttccctcaatgacatgatgatagagaactcctgca | 773 |
| Db | 661 | AAACCAAGCTGTGTCGCGGATCGAGGGCTTCCTTCAATGAGCATGTATGCAAGCTCCGACG | 720 |
| QY | 774 | cctgtaaagttagccaattccctcctaagaattgacgtctcg | 813 |
| Db | 721 | CTGGTAAAGTCCCATTTTCCTCTACGATGACCGCTTGG | 760 |

| | | | |
|------------|---|--------|-----------------|
| RESULT | 5 | | |
| B1821769 | | | |
| LOCUS | B1821769 | 955 bp | mRNA |
| DEFINITION | 603035838p1 NIH_MGC_115 Homo sapiens cDNA | linear | EST 04-OCT-2001 |
| ACCESSION | B1821769 | | |
| VERSION | B1821769 | | |
| KEYWORDS | B1821769.1 GI:15933319 | | |
| SOURCE | EST | | |
| ORGANISM | human | | |
| | Homo sapiens | | |

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|--------------------|---------|---|--|-----------------------------------|
| 1 (bases 1 to 955) | NIH-MGC | http://mgc.nci.nih.gov/ | National Institutes of Health, Mammalian Gene Collection (MGC) | Unpublished (1999) |
| | | | | Contact: Robert Strausberg, Ph.D. |

1. Issue Procurement: Life Technologies, Inc.
 2. cDNA Library Preparation: Life Technologies, Inc.
 3. DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 4. DNA Sequencing by: Incyte Genomics, Inc.
 5. Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 6. Plate: LLNL11450 Row: 1 Column: 09
 7. High quality sequence stop: 764.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .955 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5176856"
/clone_1b="NIH_MGC_115"
/lab_host="DH10b"
/notes="Organ: pooled brain, lung, testis; Vector:
pcwv-SPOrt6; site_1: NotI; site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
Oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research genetics tracking code
021. Note: this is a NIH_MGC Library."
216 a 302 c 246 g 190 t 1 others

```

| | | | | |
|---------------------------|--------|-------------------|-----------|-------------|
| Query Match | 32.88; | Score 551; | DB 10; | Length 955; |
| Best Local Similarity | 99.98; | Pred. No. 3e-270; | | |
| Matches 671; Conservative | 0; | Mismatches 0; | Indels 1; | Gaps 1; |

QY 122 agatgcccggcccgtgaccttcacccctctctgtgtgtcgcagccacctgatcc 181

[illegible]

Db 86 AAGCCACCTCTAGTCCACTGACGATTCATCTCGGCCCAAAAGTCATCAAAAGAAAGC 145
 242 TACACAAGAGCTGGAAGGACCAAGCCACAGCATCTGCAAGCAGCTGCGCTGCTCA 301
 Db 146 TACACAGAGAGCTGGAAGGACCAAGCCACAGCATCTGCAAGCAGCTGCGCTGCTCA 205
 302 GTGCATGCGGGAAGAAAGCAGCGGAGG-ATCCCTGTCTGGGCAAGCTGTGTAACAGC 360
 Db 206 GTGCATGCGGGAAGAAAGCAGCGGAGGATCTCTGTCTGGGCAAGCTGTGTAACAGC 265
 361 TCTGAAGAGATCATCTGTGCTGGAAGTCAATCAAGCTAATCAATCAATCAATCAATCA 420
 Db 266 TCTGAAGAGATCATCTGTGCTGGAAGTCAATCAAGCTAATCAATCAATCAATCAATCA 325
 421 AGCCCTGCGCAATGACCAAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCA 480
 Db 326 AGCCCTGCGCAATGACCAAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCA 385
 481 TCAACAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCAATCAATCAATCA 540
 Db 386 TCAACAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCAATCAATCAATCA 445
 541 CCACTGCGATGGAAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCAATCA 600
 Db 446 CCACTGCGATGGAAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCAATCA 505
 601 CCACTGCGATGGAAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCAATCA 660
 Db 506 CCACTGCGATGGAAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCAATCA 565
 661 TACGAAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCAATCAATCAATCA 720
 Db 566 TACGAAGAGAGCTGCTGATCAAGTCAATCAATCAATCAATCAATCAATCAATCAATCA 625
 721 TGTGTCGGTGTGTCGAGGCTTCTTCAATGAGCATGATGACAGCTCCTGCTGCTGTA 780
 Db 626 TGTGTCGGTGTGTCGAGGCTTCTTCAATGAGCATGATGACAGCTCCTGCTGCTGTA 685
 781 AGGTGCGCATTT 792
 Db 686 AGGTGCGCATTT 697
 RESULT 6
 BGS29820 826 bp mRNA linear EST 03-APR-2001
 LOCUS 6025358904F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4696752 5',
 DEFINITION mRNA sequence.
 ACCESSION BGS29820
 VERSION BGS29820.1 GI:13521357
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 826)
 AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaapb-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.llnl.gov
 Plate: L1CM1524 row: h column: 01
 High quality sequence stop: 820.
 Location/Qualifiers
 1..826
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone="IMAGE:4696752"
 /clone.lib="NIH_MGC.61"
 /tissue_type="embryonal carcinoma"
 /lab_host="DH10B (T1 phage resistant)"
 /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site:1:
 SfiI (ggcgccctggcc); Site:2: SfiI (ggcgccatggcc);
 Double-stranded cDNA was prepared from cell line RNA. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCATTTATGCC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGGCGGAGGCGCGCATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
 library."

BASE COUNT 179 a 271 c 213 g 163 t
 ORIGIN

Query Match 29.2%; Score 490; DB 10; Length 826;
 Best Local Similarity 99.8%; Pred. No. 4,5e-239;
 Matches 610; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

53 ggaagcggcggagagctccagcgtgcccaggtctgacatcctgcaactctgccccttga 112
 Db 13 ggaagcggcggagagctccagcgtgcccaggtctgacatcctgcaactctgccccttga 72
 113 caacttgggaagatggcggccgctggaacttcaaccttctctgtgtgtctggaagca 172
 Db 73 CACTGTGGGAAGATGGCGGCGCGCTGAGACTTCACTTCTGTGTGTGTGTGTGTGTGTGTGT 132
 173 ccttgcataagcgaacctcaatccatctgaagttctatctctggcccaaaagtcata 232
 Db 133 CCTTATCCAAACCCACCTCACTCCACATGAGTTCTCTATCCCGCCCAAAAGTCATCA 192
 233 aagaagaagctgacagagagctgaaaggaacacacagccacagacatcctgacagctgc 292
 Db 193 AAGAAAAGCTGACACAGAGAGCTGAAAGACACACACACACACACACACATCTGACAGCTGC 252
 293 cgtctcagtgacatgctggaagaaagccagccggaag-attccctgtgtctggaagcctg 351
 Db 253 CGCTGCTCAGTGCATGCGGGAAGCCAGCGGAGGATCCTGTGTGTGTGTGTGTGTGTGTGT 312
 352 tgaacacgctcctgaagcaatcattctgctgaaagtgatcaacagctaacatcctccagc 411
 Db 313 TGAACACCGTCTCTGAAGCACAATCATCTGCTGAAGTCAATCAAGCTAATCATCTCCACAC 372
 412 tgaagtgaaagccctcgagcaatgacaaagagctctgaatcaagatccccctggagcatg 471
 Db 373 TGAAGTGAAGCCCTCGGCAATGACCAAGAGCTGCTAAGATCCCCCTGGAGCATGG 432
 472 tggcttgatcaacaacgcccctgtgtcaagaacatcgttggagtctcaacatgacagactgaag 531
 Db 433 TGGCTGAGATTCAACAGCGCCCTGTGTCAAGACCATGCTGTGAGTTCCACATGAGACTGAAG 492
 532 cccaagcgaatcctcgatggaacacagtggaagtgagcccaacccgctgtgtcccatgt 591
 Db 493 CCCAAGCCACCATCGGCAATGACACCAAGTGAAGTGGCCGCCCGCTGTGCTCAGAG 552
 592 actgtgcaacagagcatggaagcctgagcatcaacatgctgtgataagctcctctccgtg 651
 Db 553 ACTGTGCAACGAGCATGGAAGAGCTGCGATCAACTGCTGATTAAGCTCTCTTCTCTGG 612
 652 tgaacgacctta 662
 Db 613 TGAACGCTCTTA 623

RESULT 7
 BE061388/c BE061388 658 bp mRNA linear EST 09-JUN-2000

DEFINITION QV0-BF02229-251099-038-c01 BF0229 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE061388
VERSION BE061388.1 GI:8406038
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 658)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-QV0-BF0229-251
099-038-c01&t3=1999-10-25&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 620.
Location/Qualifiers
1..658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BF0229"
/dev_stage="Adult"
/note="Organ: breast; Vector: pUC18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 138 a 161 c 214 g 144 t 1 others
ORIGIN
Query Match 28.9%; Score 485; DB 9; Length 658;
Best Local Similarity 99.7%; Pmed. No. 1.6e-236;
Matches 585; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 656 cgccttgaagcagcaggtgcatgaactctcagtcacccctccctcccaatctagtgaanaa 715
Db 322 CGCCTTAGCTAAGCAGGTCATGAAACCTCTAGGCCATCCCTGCCAATCTAGTGAAGAAA 263
QY 716 ccaagctgtcccgatgtagcagagcttccatgaatgcatgatgagagactctctgaagct 775
Db 262 CCAAGCTGTGCCGTATGATGAGGCTTCTTCAATGGCATGTATGCAAGACCTCTGCAGCT 203
QY 776 ggtgaaggtgcccattccctcagcatgtgacgclctggaagtttgactctgtatctcgc 835
Db 202 GGTGAAGGTGCCATTTCCCTCAGCATTTGACCTGTGAGATTGACCTGTGATCTCTGC 143
QY 836 catcaaggtgagacacattagcttcaactcgtggggcccaagttgtgactacacagga 895
Db 142 CATCAAGGTGAGACCATTTAGCTTACCTGGGGGCCAAGTTGTGACTCACAGGGAAA 83
QY 896 ggtgaccaggtgttcaataactctgacgttcccccgcgaatgcca 942
Db 82 GGTGACCAAGTGTTCATATACCTCTGCAGCTTCCCTGACAAATGCCA 36
RESULT 8
BF088537 567 bp mRNA linear EST 19-OCT-2000
LOCUS RCI-HT0881-110900-024-d01 HT0881 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF088537
ACCESSION BF088537.1 GI:10894247
VERSION EST.
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 567)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-RC1-HT0881-110
900-024-d01&t3=2000-09-11&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 19
High quality sequence stop: 567.
Location/Qualifiers
1..567
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0881"
/dev_stage="Adult"
/note="Organ: head,neck; Vector: pUC18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 139 a 147 c 143 g 138 t

ORIGIN

| | | | | |
|---------------------------|--------|---------------------|-----------|-------------|
| Query Match | 26.8%; | Score 451; | DB 10; | Length 567; |
| Best Local Similarity | 99.8%; | Pred. No. 3.6e-219; | | |
| Matches 501; Conservative | 0; | Mismatches 1; | Indels 0; | Gaps 0 |

| | | | |
|----|------|--|------|
| QY | 1001 | gagctgcgtgctctccccaagaattca tgg tccgtgtgacgtctggtccctcgagag | 1060 |
| Db | 53 | GGCGGTGCTCTCTCCCAAGAAATCA TGGTCCCTGTTGGACACTGTGCTTCTCGAAG | 112 |
| QY | 1061 | tgcocatcgctgaag tcaagcatcggtctga tcaatgaanaagctgcagataagcttgg | 1120 |
| Db | 113 | TGGCCATCGGCTTAACTCAGACATCGGCTGTATCAATGAAAGCGCTGCAGATAAGCTGGG | 1172 |
| QY | 1121 | atctaccagatcg taaagatccctaactcagagacatcccgagtttttatatagaacag | 1180 |
| Db | 173 | ATCTACCCAGATGTGTAAGATCTTA CTACTAGACACTCCCGAGTTTATTATAGACCAAGG | 232 |
| QY | 1181 | ccatgccaagatg tggcccaactgatcg tctgtgaag tgtttccctccag tgaagccctcg | 1240 |
| Db | 233 | CCATGCCAAAGTGGCCCAACTGATCTGTGGAGTGTTCCCTCCATGGAAGCCCTCGG | 292 |
| QY | 1241 | cccttgttcaacctgtggatcgagcca gctccggaagctca gttttacaccaaagtga | 1300 |
| Db | 293 | CCCTTGTGTACCTCTGGGATCGAAGCCAGCTGGGAAGCTCA GTTTTACCCAAAGGTGA | 352 |
| QY | 1301 | ccaactatactaa cttgaataacatcacgtctgatacgtatccagctgfatgaactctg | 1360 |
| Db | 353 | CCAACTTATTA CTGAATTAATCACTGATCGCTGTATGGAGTCCAGCTGATGAAGTCTGG | 412 |
| QY | 1361 | gattggtcgtgttccaactgtagtgc tctgtaanaaaca tcaactagatgata tcaactccat | 1420 |
| Db | 413 | GATTGGCTGTGTTCCAACTGATGTTCTGAAAAACATCACTACTGAGATCAATCCACTCAT | 472 |
| QY | 1421 | ccgtgcgtccgaac cagaatctggaataatgaatcttgggttcccaatgctatctgtgtgaagc | 1480 |
| Db | 473 | CCGTGCTGCCAACAACAAATAGGCAATTAAGATCTGGGGTCCACAGTGTATTGGTGAAGC | 532 |
| QY | 1481 | cttggagattcgaagcagctgag | 1502 |
| Db | 533 | CTTGGGATTCGAGGACGTGAG | 554 |

| FEATURES | source | 1..945 | location/Qualifiers |
|------------|--|-------------|-----------------------------|
| RESULT | 9 | | |
| LOCUS | BT772722 | 945 bp | mRNA linear EST 25-SEP-2001 |
| DEFINITION | 603053235F1 NIH_MGC_122 Homo sapiens CDNA clone IMAGE:5202801 5', | | |
| ACCESSION | BT772722 | | |
| VERSION | BT772722.1 | GI:15764300 | |
| KEYWORDS | EST. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo. | | |
| TITLE | 1 (bases 1 to 945) | | |
| JOURNAL | NIH-MGC http://mgc.nci.nih.gov/ . | | |
| COMMENT | National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L14M11508 row: e column: 10 High quality sequence stop: 856. | | |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5202801"
/clone_1id="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-Sport6,
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
BASE COUNT      208 a      306 c      243 g      188 t
ORIGIN

```

| Query Match | 26.5% | Score 446 | DB 10 | Length 945 |
|--|----------------|---------------------|----------|------------|
| Best Local Similarity | 99.7% | Preid. No. 1.5e-216 | | |
| Matches 686 | Conservative 0 | Mismatches 0 | Indels 2 | Gaps 2 |
| QY 53 ggaacgggcccgaagcactcagcctgcccaggctctgcgaatccgcgaactgctgcctctga | 112 | | | |
| DB 48 GGACCGGGCCAGAGACTCCAGCGTGCACAGTCTGGCATCTTGCACTTGCTGCTCTTGA | 107 | | | |
| QY 113 caacttggaagaatgagccgcccgttggaacttcaacccctctctgttggttctgcgaagca | 172 | | | |
| DB 108 CACCTGGGAACAATGGCCGGCCCGGTGGACCTTCACCCCTCTCTGTGTGGTTGGTGGCAGCA | 167 | | | |
| QY 173 ccttgatccaaagcgaacccctcagctcccaactgcagttctcatccctgcggccaaagtcatca | 232 | | | |
| DB 168 CCTGTGATCCAGGCCACCCCTCAGTCCCACTGCAGTGTCTCATCTCCGCCCCAAAGATCA | 227 | | | |
| QY 233 aagaaagctgaacacaggaagcttgaagagcccaaaagcacaagatctctgaagagctgc | 292 | | | |
| DB 228 AAGAAAGCTTACACACAGGAGCTGAAGGACCAACACGACCAACGACATCTGCAGCAGCTGC | 287 | | | |
| QY 293 cgcctcgaagtgcaatgctgggaaagacagccgagag-ataccctgtgtgcgaaccttg | 351 | | | |
| DB 288 CGCTGCTCAGTGGCCATGGGGGAAAAAGCCAGCCGAGGCGATCCCTGCTGGGACGCTGG | 347 | | | |
| QY 352 tgaacacccgtcctgaagcacatcat-ctggtctgaaggtcatcaacagataacatcctcag | 410 | | | |
| DB 348 TGAACACGCTCCTGAAGCACATCATCTGGCTGAAGGTGATCATCAGCTTAACATCCTCCAG | 407 | | | |
| QY 411 ctgcaggtgaaagcctctgcgccaatgacagaagctgcctcagttcaagaatcccccctgaactg | 470 | | | |
| DB 408 CTGAGAGTGAAGCCCTCGGCCAATGACCAAGGAGCGTGAATCAATATCCCTCGAGCATG | 467 | | | |
| QY 471 gtgctgattatcaacagccccctgttcaagaacatcgttggagtttccatataagcactga | 530 | | | |
| DB 468 GTGCTGTGATTCAACACGCCCTCGTGTCAAGACATCTGTGGAACTTCAATATGCGACTGG | 527 | | | |
| QY 531 gcccaagccacacatccgatgcatggaacacagttgcaatgtgcccacccgcctggtctcagt | 590 | | | |
| DB 528 GCCCAAGCCACATCCGATGGACACAGTGCATATGGCCCAACCCGCTGGTCTCTCAGT | 587 | | | |
| QY 591 gactgtgcacacagccatlggagccttcgcatcccaactgtctgacataaagctctccctcctg | 650 | | | |
| DB 588 GACTGTGCACACAGCATGGAGGACTGGGCAATCCAAATGTGTGCAATAAAGCTCTCTTCG | 647 | | | |
| QY 651 gtgaacgccttaagctaaagagtgcatgaactctcagttgacatccctgcgccaaactagt | 710 | | | |
| DB 648 GTGAACGCTTAAGCTAAACAGGTCAATGAACTCTCAAGTGCAATCCCTGCCAATCTAAGT | 707 | | | |
| QY 711 aaaaacagcgtgtgtcccttgatcagag | 738 | | | |
| DB 708 AAAAACAGCTGTGTCCGCTATCAGG | 735 | | | |


```

QY 53 ggaagcgagccgaagactcaggtgcccaggtctgcatctgacttgccctctga 112
    |||||||
Db 13 GGAGCGGGCGGAGAGACTCCAGGCTGCCAGGCTGGCATCTCTCACTTGCTGCTGTA 72
QY 113 cacttggaagatgagcgcgcgttgagacttccctctctgtgttgcttgagacca 172
    |||||||
Db 73 CACCTGGGAAGATGGCGGGCCGTGGACCTTCACCCCTTCTCTGTGGTTGCTGGAGCCA 132
QY 173 ccttgatccaagaccctctagctcccactgagttctctctctcgagcccaaaagtatca 232
    |||||||
Db 133 CCTTGATCCAAAGCCACCTTCAGTCCCACTGCACTTCATCTCCGGCCCAAAAGTCATCA 192
QY 233 aagaagaagctgacacagagagctgaagacacacagccacacagctctctgacagctgc 292
    |||||||
Db 193 AAGAAAAGCTGACACAGAGACTGAAGAGACCAACGCCACACATCTTCACAGAGCTGC 252
QY 293 cgcctgctcagtgccatgagcgggaaagccagccggaag-atacctgtgcttgagcagcttg 351
    |||||||
Db 253 CGCTGCTCAAGTGCATGGGGAAGCCAGCCGAGGCAATCCCTGTGCTGGCAGCCTGG 312
QY 352 tgaacacccgtctcgaagacacatctgtgctgaaggttcataaagctaacatctccagc 411
    |||||||
Db 313 TGAACACCGTCTGGAAGCAATCATCTGGCTGAAGGTCAACAGCTAACATCTCCACAGC 372
QY 412 tgcaggtgaagccctcgagcaatgacagagctgctagtcagagatcccccctgagacatg 471
    |||||||
Db 373 TGCAGGTGAAGCCCTTGCCCAATGACAGAGAGTCTAGTCAAGATCCCTTGGAATGG 432
QY 472 tggctgagatccaacagccctctgttcaagacacatcgltgagttccacatgaagcagtgag 531
    |||||||
Db 433 TGCTGATTCACAGCAGCCCTGCTGTAAGCAATCGTGGAGTTCCAAATGACAGCTGAGG 492
QY 532 cccaagccacacatccgcataggaacagatgagagtgcccacccgcgcctgctcaatgag 551
    |||||||
Db 493 CCCAAGCCACCAATCCGATGAGACACCAAGTGCAGTGCAGCCACCCCTGTGCTCAGTG 552
QY 592 actgtgccacacagcatggaagcctgagcctcacaactgctgataagctc 641
    |||||||
Db 553 ACTGTGCCACCAACCATGGGAGGCTGGCATTCACACTGCTGCATTAAGCTC 602

```

```

RESULT 12
BG548906 660 bp mRNA linear EST 04-APR-2001
LOCUS 602575739F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4703816 5',
DEFINITION mRNA sequence.
ACCESSION BG548906
VERSION BG548906.1 GI:13547571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 660)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LNCM1542 row: n column: 09
High quality sequence stop: 660.
Location/Qualifiers
1..660
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="IMAGE:4703816"
/clone_lib="NIH_MGC_77"
/lab_host="PH10B (T1 phage-resistant)"
/note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site:1:
Site1 (ggcgcctcgcc); Site:2: Site1 (ggcctatagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCATATGAGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGCGGCGCGCATGAG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
BASE COUNT 144 a 222 c 176 g 118 t
ORIGIN

```

```

Query Match 23.8%; Score 400; DB 10; Length 660;
Best Local Similarity 99.7%; Pred. No. 4,5e-193;
Matches 570; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

```

QY 53 ggaagcgagccgaagactcaggtgcccaggtctgcatctgacttgccctctga 112
    |||||||
Db 13 GGAGCGGGCGGAGAGACTCCAGGCTGCCAGGCTGGCATCTCTCACTTGCTGCTGTA 72
QY 113 cacttggaagatgagcgcgcgttgagacttccctctctgtgttgcttgagacca 172
    |||||||
Db 73 CACCTGGGAAGATGGCGGGCCGTGGACCTTCACCCCTTCTCTGTGGTTGCTGGAGCCA 132
QY 173 ccttgatccaagaccctctagctcccactgagttctctctctcgagcccaaaagtatca 232
    |||||||
Db 73 CACCTGGGAAGATGGCGGGCCGTGGACCTTCACCTCTCTGTGGTTGCTGGAGCCA 132
QY 173 ccttgatccaagaccctctagctcccactgagttctctctctcgagcccaaaagtatca 232
    |||||||
Db 133 CTTTGATTCACAGCAGCCCTGCTGTAAGCAATCGTGGAGTTCCAAATGACAGCTGAGG 192
QY 233 aagaagaagctgacacagagagctgaagacacacagccacacagctctgacagctgc 292
    |||||||
Db 193 AAGAAAAGCTGACACAGAGACTGAAGAGACCAACGCCACACATCTTCACAGAGCTGC 252
QY 293 cgcctgctcagtgccatgagcgggaaagccagccggaag-atacctgtgcttgagcagcttg 351
    |||||||
Db 253 CGCTGCTCAAGTGCATGGGGAAGCCAGCCGAGGCAATCCCTGTGCTGGCAGCCTGG 312
QY 352 tgaacacccgtctcgaagacacatctgctgtaaggttcataaagctaacatctccagc 411
    |||||||
Db 313 TGAACACCGTCTGGAAGCAATCATCTGGCTGAAGGTCAACAGCTAACATCTCCACAGC 372
QY 412 tgcaggtgaagccctcgagcaatgacagagctgctagtcagagatcccccctgagacatg 471
    |||||||
Db 373 TGCAGGTGAAGCCCTTGCCCAATGACAGAGGCTGCTAGTCAAGATCCCTTGGAATGG 432
QY 472 tggctgagatccaacagccctctgttcaagacacatcgltgagttccacatgaagcagtgag 531
    |||||||
Db 433 TGCTGATTCACAGCAGCCCTGCTGTAAGCAATCGTGGAGTTCCAAATGACAGCTGAGG 492
QY 532 cccaagccacacatccgcataggaacagatgagagtgcccacccgcgcctgctcaatgag 551
    |||||||
Db 493 CCCAAGCCACCAATCCGATGAGACACCAAGTGCAGTGCAGCCACCCCTGTGCTCAGTG 552
QY 592 actgtgccacacagcatggaagcctgagcctcacaactgctgataagctc 641
    |||||||
Db 553 ACTGTGCCACCAACCATGGGAGGCTGGCATTCACACTGCTGCATTAAGCTC 602

```

```

RESULT 13
BE933204 524 bp mRNA linear EST 02-OCT-2000
LOCUS R01-HT0881-290800-022-e03 HT0881 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE933204
ACCESSION BE933204
VERSION BE933204.1 GI:10459280
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 524)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Joergensen,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 JOURNAL MEDLINE
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LIOR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=62-rc1-HT0861-290800-022-e03&t3=2000-08-29&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 13
 High quality sequence stop: 524.

FEATURES
 source
 1..524
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="HT0861"
 /dev_stage="Adult"
 /note="Organ: head neck; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 127 a 145 c 130 g 122 t
 ORIGIN

Query Match 23.6%; Score 396; DB 10; Length 524;
 Best Local Similarity 99.8%; Pred. No. 4.8e-191;
 Matches 446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 911 caataactctgacgtccctctgacaaagtcgacacacatccgcttcagcctcat 970
 Db 78 CAATTAACCTCTCAGCTTCCCTGACAAATGCCACCTGGACAACATCCCGTTACAGCTTCAT 137
 QY 971 cgtgaagtcagacgtctggaagtcgacgtctgctctctcagaagaattcat 1030
 Db 138 CGTGAGTGAAGACGTGGTGAAGAGCTGCAGTGGCTGTCTCTCTCCAGAAATTCAT 197
 QY 1031 gctcctgtgactctgtctctctgagagtcgcccacatcgctggaagtcagcggct 1090
 Db 198 GGTCTCTGTGACCTCTGCTCTCCGAGAGTGCCTGCGTGAAGTCAACATCGGCGCT 257
 QY 1091 gatacaataaagctgtagaagtcgtgatatcccaagtcgtgaagatccctaacc 1150
 Db 258 GATCAATGGAAGGCTGAGATTAACCTGGATCTACCCAGATCGGAAGATCTTAATCA 317
 QY 1151 ggacaccccgagtttttatagaccagagccatcccaagtcggccacatgactgtct 1210
 Db 318 GGACACTCCCGAGTTTATATAGACCAAGCGCATCGCAAGGTGGCCCAACTGATGTGCT 377
 QY 1211 ggaagtgttccctccacagtgaagccctcgcaccttgttcaacctgggcatcgagccag 1270
 Db 378 GGAAGTGTTCCTCCCTCAGTGAAGCCCTCGCCCTTTGTTCAACCCCTGGCATCGAAGCCAG 437
 QY 1271 ctgggaagctcaagttttacaccaagtgagcaacttatctcaacttgaataacatcag 1330
 |||||||

Db 438 CTGGAAGCTCACTTTTACACCAAGGTGACCACTTATCTCACTTGAAATACATCAG 497
 QY 1331 cctgatacgatccagctgatgacac 1357
 |||||||
 Db 498 CTCTGATCGGATCGACTGATGACAC 524

RESULT 14
 BI824243 539 bp mRNA linear EST 04-OCT-2001
 LOCUS 603040625F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5181222 5',
 DEFINITION mRNA sequence.
 ACCESSION BI824243
 VERSION BI824243.1 GI:15935793
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 539)
 AUTHORS NIH-MGC http://img.ncbi.nlm.nih.gov/.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: LLAM1452 row: b column: 07
 High quality sequence stop: 535.

FEATURES
 source
 1..539
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="IMAGE:5181222"
 /clone_1lb="NIH_MGC_115"
 /lab_host="DH10B"
 /note="Organ: pooled brain, lung, testis; Vector: pCMV-Sport6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research genetics tracking code 021. Note: this is a NIH-MGC library."

BASE COUNT 119 a 184 c 141 g 94 t 1 others
 ORIGIN

Query Match 23.3%; Score 392; DB 10; Length 539;
 Best Local Similarity 99.8%; Pred. No. 5.3e-189;
 Matches 512; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 62 cgaagactccagcgtgcccaggtctgacatccctcactgctgcgcctctgacacactgga 121
 Db 7 CGAGAGCTCCAGCGTCCGACAGTCTGGCATCTGCTGCTGCTGCGCATCTGACACTGGA 66
 QY 122 agatggccggccgttgacacttccacctctctgtgtgtgtgtgagcagcaactgtatcc 181
 Db 67 AGATGGCCGGCCGCTGGAGCTTCAACCTTCTGTGCTTGTGCTGGAGCCACCTTGATTC 126
 QY 182 aagccacctatgctccacatgcaattctcatccctcggcccaaaagtaataaagaagc 241
 Db 127 AAGCCACCTCTAGTCCACATGCACTTCATCTCGGCGCCCAAAAGTATCAAAAGAAAGC 186
 QY 242 tgacaagagctgaagacacacagccacacacatccctcagcagctcggctgtctca 301
 |||||||

```

Db      187 TGACACAGAGAGCTGAAGAGCACACACAGCAGCATCTCGAGAGAGCTGCCGTCTCA 246
Qy      302 gtgcacatgggggaaagccagccgagag--atccctgtgtgggagccgtgtgaacacg 360
Db      247 GTGCCATGGGGGAAAGCCAGCCGGAGGAGCATCCCTGTGCTGGGACAGCTGGTGAACACCG 306
Qy      361 tcttgaagacatcatctgtctgaaggtatcatcacagctaacatccctcagctgagtgat 420
Db      307 TCCGTGAAGACATCATCTGGCTGAAGGTCATCACACCTAACATCCCTCCAGAGTCAGAGTCA 366
Qy      421 agccctgggccaatgacagagagctgtctgaagaatccccctgagacatgtgtgctgat 480
Db      367 ACCCCCTGGGCCAATGACCGAGAGCTGCTAGTCAAGATCCCTCGACATGTGGCTGAGT 426
Qy      481 tcaaacgcccctgtctgaagacatctgtgaagttccacatgagcagactgagagccca 540
Db      427 TCAACACGCCCTGTGTCAGAGACCATGTGTGAGTCCACATGACGACTGAGGCCCAAGCA 486
Qy      541 ccacccgcatggacacacagtgcaagtggcccca 573
Db      487 CCATCCGATGGACACCACTGCAAGTGGCCCA 519

RESULT 15
BF154920      421 bp      mRNA      linear      EST 30-OCT-2000
LOCUS      RC0-BT0812-250900-032-e11 BT0812 Homo sapiens cDNA, mRNA sequence.
DEFINITION      BF154920
ACCESSION      BF154920.1 GI:11050103
VERSION      EST.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 421)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
      Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
      Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
      Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
      ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
      Simpson,A.J.
      Shotgun sequencing of the human transcriptome with ORF expressed
      sequence tags
      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
      Contact: Simpson A.J.G.
      Laboratory of Cancer Genetics
      Ludwig Institute for Cancer Research
      Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
      Brazil
      Tel: +55-11-2704922
      Fax: +55-11-2707001
      Email: asimpson@ludwig.org.br
      This sequence was derived from the FAPESP/LICR Human Cancer Genome
      Project. This entry can be seen in the following URL
      (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=6r2-RC0-BT0812-250
      900-032-e11&f3=2000-09-25&f4=1)
      Seq primer: puc 18 forward
      High quality sequence start: 3
      High quality sequence stop: 421.
      Location/Qualifiers
      1..421
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="BT0812"
      /dev_stage="Adult"
      /note="Organ: breast; Vector: puc18; Site:1: SmaI; Site:2:
      SmaI; A mini-library was made by cloning products derived
      from ORESTES PCR (U.S. Letters Patent application No. 196
      ,716 - Ludwig Institute for Cancer Research) profiles
      into the pUC 18 vector. Reverse transcription of tissue
      mRNA and cDNA amplification were performed under low
      stringency conditions."

```

```

BASE COUNT      99 a      115 c      107 g      100 t
ORIGIN
Query Match      23.2%; Score 390; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 5.3e-188;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      818 tgacctctgttccctgcacataaagggtgacccattcagctctactctggggccaagt 877
Db      32 TGACCTTGTATCTTCGCATCAAGGGTGAACACATTCAGCTCTACCTGGGGCCAAAGTT 91
Qy      878 gttagacacagggaaagtgacaaagtgttcaactctgacgtccctcgaat 937
Db      92 GTTGGACTACAGAGGAAAGTGAACAGTGGTTCAATTAATCTGCGAGCTTCCCTGCAAT 151
Qy      938 gcccaacctggacaacacatcccgltcagctcactcgtgagtcagagctggtgaaagctgc 997
Db      152 GCCACCCCTGGACAAACATCCCGTTACAGCTCATCGTAGTCAAGAGAGTGTAAGACTGC 211
Qy      998 agtggctgtgtgtctcttccagaagaattcaatggtccctgttgactctgtgttcctga 1057
Db      212 AGTGGCTGTGTCTCTCTCCAGAGAAATTCATGATGTCGTGGACTCTGTGCTCTCTGA 271
Qy      1058 gaatgccatcggtctgaagcaatcgagctgacatgaaagctgcagataagct 1117
Db      272 GAGTGGCCATCGGCTGAAGTCAGATCGGGCTGATCANTGAAGGCTGCAGATAGCT 331
Qy      1118 ggaattacacagatcgtgaagatccttaactcaaggaactccgagttttatagaca 1177
Db      332 GGGATCTACCCAGATCGTAGAGATCCTTAATCTCAGCACTCCCGAGTTTATAGACA 391
Qy      1178 aggcacatgccaaagtgtgcccactgaatcgt 1207
Db      392 AGGCCATGCCAAGGTGGCCCAACTGATCGI 421

```

Search completed: September 16, 2002, 05:18:12
Job time: 8385 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 00:52:02 : Search time 390.12 seconds
(without alignments)
7393.664 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 ggtgtgcagagataaggtt.....tcaataaaccttgctgtg 1680

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq.032802:*

1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|----------|---------------------|
| 1 | 1679.6 | 100.0 | 1680 | 21 | AAZ29724 | Human Lung specific |
| 2 | 1679.2 | 100.0 | 1680 | 22 | AAH77948 | Nucleotide sequenc |
| 3 | 1667 | 99.2 | 1707 | 21 | AAH87210 | Human NTAP CDNA c1 |
| 4 | 1656.2 | 98.6 | 2036 | 22 | AAH77950 | Nucleotide sequenc |
| 5 | 1617 | 96.2 | 1636 | 21 | AAA37057 | Human PRO1357 (UNO |
| 6 | 1617 | 96.2 | 1636 | 22 | AA546064 | Human DNA encoding |
| 7 | 1617 | 96.2 | 1636 | 22 | AAH92096 | Human PRO1357 CDNA |
| 8 | 1617 | 96.2 | 1636 | 22 | AAH54293 | DNA encoding prote |
| 9 | 1533 | 91.2 | 2121 | 23 | AA572707 | DNA encoding novel |

| | | | | | | |
|----|-------|------|------|----|----------|--------------------|
| 10 | 806 | 48.0 | 1035 | 22 | ABA01989 | Human NOV6b coding |
| 11 | 724.2 | 43.1 | 963 | 22 | ABA01986 | Human NOV6a coding |
| 12 | 641 | 38.2 | 1172 | 21 | AAZ97114 | Human secreted pro |
| 13 | 635.8 | 37.8 | 1177 | 21 | AAH18254 | Lung cancer associ |
| 14 | 634.8 | 37.8 | 1175 | 21 | AAZ97023 | Human secreted pro |
| 15 | 420 | 25.0 | 420 | 21 | AAH30554 | Human colon cancer |
| 16 | 417.4 | 24.8 | 465 | 21 | AAZ74959 | Human OREF514 |
| 17 | 398.2 | 23.7 | 423 | 22 | AAH19484 | Human breast cance |
| 18 | 373.2 | 22.2 | 462 | 22 | AAH09812 | Human breast cance |
| 19 | 229 | 13.6 | 232 | 16 | AAH21378 | Human gene signatu |
| 20 | 181.4 | 10.8 | 594 | 22 | ABA61288 | Human foetal liver |
| 21 | 181.4 | 10.8 | 594 | 22 | ABA29112 | Probe #7578 for ge |
| 22 | 181.4 | 10.8 | 594 | 22 | AAK09583 | Human brain expres |
| 23 | 181.4 | 10.8 | 594 | 22 | AAK35477 | Human bone marrow |
| 24 | 181.4 | 10.8 | 594 | 22 | AAH17050 | Probe #6983 for ge |
| 25 | 181.4 | 10.8 | 594 | 22 | AAH41191 | Probe #9877 used t |
| 26 | 180 | 10.7 | 180 | 22 | ABA73788 | Human foetal liver |
| 27 | 180 | 10.7 | 180 | 22 | ABA38963 | Probe #17429 for g |
| 28 | 180 | 10.7 | 180 | 22 | AAK22236 | Human brain expres |
| 29 | 180 | 10.7 | 180 | 22 | AAK48404 | Human bone marrow |
| 30 | 180 | 10.7 | 180 | 22 | AAH26228 | Probe #16161 for g |
| 31 | 180 | 10.7 | 180 | 22 | AAH54231 | Probe #22917 used |
| 32 | 163.4 | 9.7 | 576 | 21 | AAH74887 | Human OREF ORF42 |
| 33 | 161 | 9.6 | 583 | 22 | ABA60344 | Human foetal liver |
| 34 | 161 | 9.6 | 583 | 22 | ABA28591 | Probe #7057 for ge |
| 35 | 161 | 9.6 | 583 | 22 | AAK08624 | Human brain expres |
| 36 | 161 | 9.6 | 583 | 22 | AAK34508 | Human bone marrow |
| 37 | 161 | 9.6 | 583 | 22 | AAH16752 | Probe #6685 for ge |
| 38 | 161 | 9.6 | 583 | 22 | AAH40226 | Probe #8912 used t |
| 39 | 156 | 9.3 | 156 | 22 | ABA72883 | Human foetal liver |
| 40 | 156 | 9.3 | 156 | 22 | ABA38468 | Probe #16934 for g |
| 41 | 156 | 9.3 | 156 | 22 | AAK23116 | Human brain expres |
| 42 | 156 | 9.3 | 156 | 22 | AAK47474 | Human bone marrow |
| 43 | 156 | 9.3 | 156 | 22 | AAH25947 | Probe #13880 for g |
| 44 | 156 | 9.3 | 156 | 22 | AAH53505 | Probe #21991 used |
| 45 | 141 | 8.4 | 375 | 23 | AA572706 | DNA encoding novel |

ALIGNMENTS

RESULT 1

AAZ29724

AAZ29724 standard; DNA; 1680 BP.

AAZ29724:

27-MAR-2000 (first entry)

Human Lung specific gene-1.

Human Lung specific Gene; LSG; human; diagnostic marker;

prognosticate; Lung cancer; diagnosis; ds.

Homo sapiens.

WO9960160-A1.

25-NOV-1999.

12-MAY-1999; 99WO-US10344.

21-MAY-1998; 98US-0086212.

(DIAD-) DIADEXUS LLC.

Yang F, Macina RA, Sun Y;

WPI; 2000-116320/10.

A new method for diagnosing, monitoring and staging lung cancer

Example 1; Pages 33-34; 40pp; English.


```
FT      " (pos: 328..329, aa: Gly) "
```

```

/ transl_except= " (pos: 372..374, aa: Val) "

```

W0200161055-A2.

23-AUG-2001.

20-FEB-2001; 2001WO-US05674.

17-FEB-2000; 2000US-0183188.

(DIAD-) DIADEXUS INC.

Chen S, Sun Y, Macina RA;

WPI; 2001-529917/58.

P-PSDB; AAG63976.

New lung cancer specific gene for the treatment and diagnosis of lung cancer -

Claim 1; Page 112-113; 119pp; English.

The present sequence represents a human lung cancer specific gene (LSG) and encodes a polypeptide designated Lng103. LSGs are useful in the treatment and diagnosis of lung cancer. The treatment of lung cancer comprises the administration of a molecule which down regulates the expression of an LSG. An immune response can be mounted against a target cell expressing an LSG. Identification of potential therapeutic agents for use in imaging and treating lung cancer which comprises screening molecules for an ability to bind to or decrease expression of an LSG relative to LSG in the absence of the agent where the ability of a molecule to bind to the LSG or decrease expression of the LSG is indicative of the molecule being useful in imaging and treating lung cancer.

SQ Sequence 1680 BP; 380 A; 503 C; 429 G; 367 T; 1 other;

| Query Match | 100.08; | Score 1679.2; | DB 22; | Length 1680; |
|-------------|---------|---------------|--------|--------------|
|-------------|---------|---------------|--------|--------------|

Best Local Similarity 99.98; Pred. No. 0;

Matches 1678; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 1 | ggtctgcaggaatataaaggttgtagacttccaaacacactgcccgggagaagagatgagccggg | 60 |
| Db | 1 | ggctctgcaggaatataaaggttgtagacttccaaacacactgcccgggagaagagatgagccggg | 60 |
| QY | 61 | ccagagagctccaaagctgtcccaagtlctgcatactctgaactgtgcgcctctgcagaaacttggg | 120 |
| Db | 61 | ccagagagctccaaagctgtcccaagtlctgcatactctgaactgtgcgcctctgcagaaacttggg | 120 |
| QY | 121 | aagaatggccggccggttgcaccttcaacctctctctgtgattgtctgcgcagccaaactgtatc | 180 |
| Db | 121 | aagaatggccggccggttgcaccttcaacctctctctgtgattgtctgcgcagccaaactgtatc | 180 |
| QY | 181 | caagaccacccctcagtlcccaactgtcagttctcatcctctcgcccaaaagtctacatcaagaagaag | 240 |
| Db | 181 | caagaccacccctcagtlcccaactgtcagttctcatcctctcgcccaaaagtctacatcaagaagaag | 240 |
| QY | 241 | ctgcagcaaggaagcttgcgaagagacacaaagccaaagcatcctctgcagcaagcttgcgcgtgttc | 300 |
| Db | 241 | ctgcagcaaggaagcttgcgaagagacacaaagccaaagcatcctctgcagcaagcttgcgcgtgttc | 300 |
| QY | 301 | agtgcacatggcggaagaagccagccggcgaggatccctctgtgcgtggcgaagcctgtgttaaacgg | 360 |
| Db | 301 | agtgcacatggcggaagaagccagccggcgaggatccctctgtgcgtggcgaagcctgtgttaaacgg | 360 |
| QY | 361 | tctcgaagcacatcatctgtgctgaaggtctcaacagctaaacatccctccagctcaggtta | 420 |
| Db | 361 | tctcgaagcacatcatctgtgctgaaggtctcaacagctaaacatccctccagctcaggtta | 420 |
| QY | 421 | agcccttcggccaatgacccaggaagctgtgattcaagaatcccccctgcagatgtgtgcgcggt | 480 |
| Db | 421 | agcccttcggccaatgacccaggaagctgtgattcaagaatcccccctgcagatgtgtgcgcggt | 480 |

| | | | | |
|---|---|------|---|------|
| D | b | 421 | agcccttcgycgaatctgacccagcagtcgtctatgctcaagaatccccccttgtaacatagtgtgctgtagt | 480 |
| Q | y | 481 | tcaacacgccccctggtctcaagaaaccatctgtgagttctccacaatgacagactagagcccaagcaca | 540 |
| D | b | 481 | tcaacacgccccctggtctcaagaaaccatctgtgagttctccacaatgacagactagagcccaagcaca | 540 |
| Q | y | 541 | ccatccgcatactgacacccaagtctgacagctgagccccaaccccgcttgcttcagtgactgtctgcca | 600 |
| D | b | 541 | ccatccgcatactgacacccaagtctgacagctgagccccaaccccgcttgcttcagtgactgtctgcca | 600 |
| Q | y | 601 | ccagcccaatggagaccctgagcagatcccaactgtctcacaagaattcccttccttccttggtgaacgcct | 660 |
| D | b | 601 | ccagcccaatggagaccctgagcagatcccaactgtctcacaagaattcccttccttccttggtgaacgcct | 660 |
| Q | y | 661 | taagtaagcagcgtctcagaaactctctctagctgccaatcccttcgcccacatctagtgaaaaaccagc | 720 |
| D | b | 661 | taagtaagcagcgtctcagaaactctctctagctgccaatcccttcgcccacatctagtgaaaaaccagc | 720 |
| Q | y | 721 | tgatgcgccgtgatccgagagcttccttcacatgtgcaatgtatgagagacactccctgcagactgtgtga | 780 |
| D | b | 721 | tgatgcgccgtgatccgagagcttccttcacatgtgcaatgtatgagagacactccctgcagactgtgtga | 780 |
| Q | y | 781 | agatgagcccaatttccttcacagatattgaaacgcctctgagatgttgaaactctctgtatccctgacatca | 840 |
| D | b | 781 | agatgagcccaatttccttcacagatattgaaacgcctctgagatgttgaaactctctgtatccctgacatca | 840 |
| Q | y | 841 | agagtgacacacattccacgctctacactctgagggcccaagtgtgttgacttcacacagaggaagaagtga | 900 |
| D | b | 841 | agagtgacacacattccacgctctacactctgagggcccaagtgtgttgacttcacacagaggaagaagtga | 900 |
| Q | y | 901 | ccaaagtgtctcaataacactctgcagactccctccctgacaaatgcacaccccttggaacaaatcccgct | 960 |
| D | b | 901 | ccaaagtgtctcaataacactctgcagactccctccctgacaaatgcacaccccttggaacaaatcccgct | 960 |
| Q | y | 961 | tcagagcccatcctgagatgcagaagaagctgtgataaagctgcagatggactctgtgactctctccag | 1020 |
| D | b | 961 | tcagagcccatcctgagatgcagaagaagctgtgataaagctgcagatggactctgtgactctctccag | 1020 |
| Q | y | 1021 | aagaaattcagctgcctcgtcttgagactctcgtctctccctcctcctcctcctcctcctcctcctcctcctc | 1080 |
| D | b | 1021 | aagaaattcagctgcctcgtcttgagactctcgtctctccctcctcctcctcctcctcctcctcctcctcctc | 1080 |
| Q | y | 1081 | gcatccgagctgcataatgtgaaaagagctgcagataagcttgagatctaacccagatctgtgaaga | 1140 |
| D | b | 1081 | gcatccgagctgcataatgtgaaaagagctgcagataagcttgagatctaacccagatctgtgtgaaga | 1140 |
| Q | y | 1141 | tcctaacatcagagacactcccgagttcttttatatagacaaagccaatgcagaaagtgtgcccac | 1200 |
| D | b | 1141 | tcctaacatcagagacactcccgagttcttttatatagacaaagccaatgcagaaagtgtgcccac | 1200 |
| Q | y | 1201 | tgatctgctgcaggaagtgtttcccttcacagtgaaagccctccgccccttggtctcaacccctgggcca | 1260 |
| D | b | 1201 | tgatctgctgcaggaagtgtttcccttcacagtgaaagccctccgccccttggtctcaacccctgggcca | 1260 |
| Q | y | 1261 | tcgaagcagcagctcgggaagctcagtttctaacacaaagtgacaaacttatatactcaactctga | 1320 |
| D | b | 1261 | tcgaagcagcagctcgggaagctcagtttctaacacaaagtgacaaacttatatactcaactctga | 1320 |
| Q | y | 1321 | ataacatcagctctgtatctcgatctccagctgataactctgagatctgagctgtgtcttccaaactg | 1380 |
| D | b | 1321 | ataacatcagctctgtatctcgatctccagctgataactctgagatctgagctgtgtgttccaaactg | 1380 |
| Q | y | 1381 | atgtgttcgaaaaaacaatcatcatcagtaatacatcatccatccatctctgtctgcgcgaacccaagatg | 1440 |
| D | b | 1381 | atgtgttcgaaaaaacaatcatcatcagtaatacatcatccatccatctctgtctgcgcgaacccaagatg | 1440 |
| Q | y | 1441 | gcaaatataagatctgaggtctccagctgtcaattgtgtaagagccttgaggatctcgagagcagctg | 1500 |
| D | b | 1441 | gcaaatataagatctgaggtctccagctgtcaattgtgtaagagccttgaggatctcgagagcagctg | 1500 |
| Q | y | 1501 | agtcctcaatcagcacaagagatgcacctgtgtactctccacagctctcctgtgtgaaacccaagct | 1560 |
| D | b | 1501 | agtcctcaatcagcacaagagatgcacctgtgtactctccacagctctcctgtgtgtgaaacccaagct | 1560 |

QY 1561 ctctctgtccaccagtgagacttgatggcagccatcaaggagagtggtgtccacgtg 1620
 |||||||
 Db 1561 ctctctgtccaccagtgagacttgatggcagccatcaaggagagtggtgtccacgtg 1620
 QY 1621 gagtatgggtgtgagctctatagacacccctctctgtcaataaataaactgtg 1680
 |||||||
 Db 1621 gegtatgggtgtgagctctatagacacccctctctgtcaataaataaactgtg 1680

RESULT 3
 AA287210
 ID AA287210 standard; cDNA, 1707 BP.
 XX
 AC AA287210;
 XX
 DT 08-MAY-2000 (first entry)
 XX
 DE Human NTAP cDNA clone 2799056.
 XX
 KW Neurotransmission-associated protein; NTAP; odorant-binding protein;
 KW neurological disease; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; cerebral neoplasm; multiple sclerosis;
 KW drug screening; gene therapy; antagonist; cancer; AIDS; asthma;
 KW Crohn's disease; osteoporosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FH CDS 124..1578
 FT /*tag= a
 FT /product= "Human neurotransmission-associated protein
 FT (NTAP) 2799056"
 XX
 FN WO200001821-A2.
 XX
 PD 13-JAN-2000.
 PD
 PE 02-JUL-1999; 99WO-US15121.
 XX
 PR 02-JUL-1998; 98US-0091677.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Lal P, Tang YT, Yue H, Corley NC, Guegler KJ, Gorgone GA;
 PI Baughn MR, Patterson C;
 XX
 DR WPI: 2000-160770/14.
 DR P-PSDB; AAY77126.
 XX
 PT New human neurotransmission associated proteins, useful for treatment,
 PT prevention and diagnosis of neurological disease, e.g. Alzheimer's
 PT disease, and antagonists for treating cancer or immune disorders -
 XX
 PS Claim 7, Page 67; 67pp: English.
 XX
 CC Sequences AA287205-287210 represent cDNAs encoding six human
 CC neurotransmission-associated proteins (NTAPs, AAY77121-Y77126). The
 CC present sequence, clone 2799056, encodes a putative odorant-binding
 CC protein, and was produced by extension of cDNA fragments isolated from a
 CC human nasal polyp tissue cDNA library. The NTAPs are used for treatment
 CC or prevention of neurological diseases (e.g., Alzheimer's, Parkinson's
 CC or Huntington's diseases, cerebral neoplasms, or multiple sclerosis).
 CC They can also be used to raise specific antibodies and to screen for
 CC specific binding agents (potential (agonists and antagonists).
 CC NTAP-encoding nucleic acids are useful for recombinant production of
 CC NTAPs, and as a source of therapeutic antagonists (antisease,
 CC triplex-forming or ribozyme molecules). The nucleic acids may also be
 CC used as a source of probes and primers for diagnosis or monitoring of
 CC NTAP expression in hybridisation/amplification tests, for chromosome
 CC mapping and for identifying related sequences, and for gene therapy. NTAP
 CC antagonists are used to treat and prevent a wide range of cancers and
 CC immune disorders (e.g. AIDS, asthma, Crohn's disease, osteoporosis, viral

CC or other infections). NTAP antibodies are used to detect NTAPs, for
 CC diagnosis or monitoring, as therapeutic antagonists, in competitive drug
 CC screens, and for affinity purification of NTAPs from natural sources.
 XX
 S0 Sequence 1707 BP; 406 A; 503 C; 429 G; 369 T; 0 other;

Query Match 99.2%; Score 1667; DB 21; Length 1707;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1678; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 ggttgcaagatataaaggttggaattccagaccacccagtcggcgagagagcgcg 60
 |||||||
 Db 1 ggttgcaagatataaaggttggaattccagaccacccagtcggcgagagagcgcg 60
 QY 61 ccgaggaactcgaagctgagccaggtgacatctctgactgtgacccctgacacccg 120
 |||||||
 Db 61 ccgaggaactcgaagctgagccaggtgacatctctgactgtgacccctgacacccg 120
 QY 121 aagatggccggcccggtgacacttcaacctctctgtgtgttgctggagaccattgac 180
 |||||||
 Db 121 aagatggccggcccggtgacacttcaacctctctgtgtgttgctggagaccattgac 180
 QY 181 caagccacccctcaagttcccaactgcaattctcatctctcgcccaaaagtcataaagaag 240
 |||||||
 Db 181 caagccacccctcaagttcccaactgcaattctcatctctcgcccaaaagtcataaagaag 240
 QY 241 ctgacacagagagctgaaagacacacacgacacacagatctctgagagctgacgctgctc 300
 |||||||
 Db 241 ctgacacagagagctgaaagacacacacgacacacagatctctgagagctgacgctgctc 300
 QY 301 agtgcacatggggaaagaaagccagccgagag- atccctgtgctggagagcctgtgtgaacc 359
 |||||||
 Db 301 agtgcacatggggaaagaaagccagccgagagatccctgtgctggagagcctgtgtgaacc 359
 QY 360 gtccctgaagacacatcatctgtgctgaaggtcatcacagacacacacacacacacacac 419
 |||||||
 Db 361 gtccctgaagacacatcatctgtgctgaaggtcatcacagacacacacacacacacacac 419
 QY 420 aagccctggcccaatgacacagagagctgtaagtcataagatcccccgtgacatggtgtg 479
 |||||||
 Db 421 aagccctggcccaatgacacagagagctgtaagtcataagatcccccgtgacatggtgtg 479
 QY 480 ttcaacacgccccctgtgtcaagacacatcgtgtgaagttccacatgacagacagccaaagcc 539
 |||||||
 Db 481 ttcaacacgccccctgtgtcaagacacatcgtgtgaagttccacatgacagacagccaaagcc 539
 QY 540 accatccgcatggaacacacagctgcaagttgcccacacccgctgtgctcaagtactgtg 599
 |||||||
 Db 541 accatccgcatggaacacacagctgcaagttgcccacacccgctgtgctcaagtactgtg 599
 QY 600 accagcagtgagagcctgtgcatccaactgctgacataagctctctctgtgtgaagcc 659
 |||||||
 Db 601 accagcagtgagagcctgtgcatccaactgctgacataagctctctctgtgtgaagcc 659
 QY 660 ttagctaaagcagtgatgaacctctagtgccatccctgtgcccacatctagtgaacacag 719
 |||||||
 Db 661 ttagctaaagcagtgatgaacctctagtgccatccctgtgcccacatctagtgaacacag 719
 QY 720 ctgtgtcccggtatgagagcttctctcaatgcatgtatgcaagacctctgacagctgtg 779
 |||||||
 Db 721 ctgtgtcccggtatgagagcttctctcaatgcatgtatgcaagacctctgacagctgtg 779
 QY 780 aagtgccacattccctcctagacatgacgctgaggttgagcttctgtatccctggcacc 839
 |||||||
 Db 781 aagtgccacattccctcctagacatgacgctgaggttgagcttctgtatccctggcacc 839
 QY 840 aaggtgacacacatcagctctactctggtggcgaagtgtgtgacacacaggaaggtg 899
 |||||||
 Db 841 aaggtgacacacatcagctctactctggtggcgaagtgtgtgacacacaggaaggtg 899
 QY 900 accaagtggttcaataaactctgacgttccctgacacatgcccacccctggaacaacatccg 959
 |||||||

| Db | 901 | accgaatggttcaataactctctgacgtcttccctcgacaatgcccacccttgacaacatcccg | 960 |
|----------|---|---|------|
| QY | 960 | tcaagccctcacctctgagtcagagagctggtgtaaaagctgcagctggtcgtgctctctcca | 1019 |
| Db | 961 | ttcagctctacgttgagtcagagagctggtgtaaaagctgcagctggtcgtgctctctcca | 1020 |
| QY | 1020 | gaagaaatcaatggtctctgttgagactctgtgtcttctctgagagatgcccacctggtgagtc | 1079 |
| Db | 1021 | gaagaaatcaatggtctctgttgagactctgtgtcttctctgagagatgcccacctggtgagtc | 1080 |
| QY | 1080 | agcaatcggtgtgttcataatgaaagagctgtaagataagctgtgagatctaacccagatcgtgag | 1139 |
| Db | 1081 | agcaatcggtgtgttcataatgaaagagctgtaagataagctgtgagatctaacccagatcgtgag | 1140 |
| QY | 1140 | atccctaactaagacacctcccaggtttcttataagaccaagagcagatgagcagatgtgccc | 1199 |
| Db | 1141 | atccctaactaagacacctcccaggtttcttataagaccaagagcagatgagcagatgtgccc | 1200 |
| QY | 1200 | ctgacgtgctgtagaagtggttccctccagtgaaagccctccgaccttgttcaacctgtgagc | 1259 |
| Db | 1201 | ctgacgtgctgtagaagtggttccctccagtgaaagccctccgaccttgttcaacctgtgagc | 1260 |
| QY | 1260 | atcgaaaccagagctcggaagctcaagttttacaccaaaagtgacacaactatactcaactg | 1319 |
| Db | 1261 | atcgaaaccagagctcggaagctcaagttttacaccaaaagtgacacaactatactcaactg | 1320 |
| QY | 1320 | aataacatcagctctgatactcgatccagctgcagatgcagactctgagattggtcgtgttccaact | 1379 |
| Db | 1321 | aataacatcagctctgatactcgatccagctgcagatgcagactctgagattggtcgtgttccaact | 1380 |
| QY | 1380 | gattgttctgaaaaacatcatcatctgagatccatccactccatctgcttgcgaaaccagat | 1439 |
| Db | 1381 | gattgttctgaaaaacatcatcatctgagatccatccactccatctgcttgcgaaaccagat | 1440 |
| QY | 1440 | ggcacaatataagatctggtgtcccaagtgtcaattggtgaaagagctctgagattcgaagcagct | 1499 |
| Db | 1441 | ggcacaatataagatctggtgtcccaagtgtcaattggtgaaagagctctgagattcgaagcagct | 1500 |
| QY | 1500 | gagttcccaatcagacaaagatgacctgtgcttactccagacctctcttcttgaaaccagc | 1559 |
| Db | 1501 | gagttcccaatcagacaaagatgacctgtgcttactccagacctctcttcttgaaaccagc | 1560 |
| QY | 1560 | tctctcgtctcccaagtgaagacttgatgtgcagacatcagaggaagagctctggttcccaagt | 1619 |
| Db | 1561 | tctctcgtctcccaagtgaagacttgatgtgcagacatcagaggaagagctctggttcccaagt | 1620 |
| QY | 1620 | gagagataggggtgtgagctctatagaccatccctctctgcaatcaataaacactgtcgt | 1679 |
| Db | 1621 | gagagataggggtgtgagctctatagaccatccctctctgcaatcaataaacactgtcgt | 1680 |
| QY | 1680 | g | 1680 |
| Db | 1681 | g | 1681 |
| RESULT 4 | | | |
| AAH77950 | | | |
| ID | AAH77950 standard; DNA; 2036 BP. | | |
| XX | AAH77950. | | |
| AC | 13-NOV-2001 (first entry) | | |
| DT | Nucleotide sequence of a human lng103 polypeptide. | | |
| XX | Human; lung cancer specific gene; LSG; lng103; lung cancer; ss. | | |
| KW | Homo sapiens. | | |
| XX | Key | | |
| FT | Location/Qualifiers | | |
| CD | 163..1617 | | |
| FT | /tag= a | | |
| TT | /product= "lng103" | | |

| | |
|---|---|
| FT | /transl_except= "(pos: 412..414, aa: Val)" |
| XX | |
| PN | WO200161055-A2. |
| XX | |
| PD | 23-AUG-2001. |
| XX | |
| PF | 20-FEB-2001; 2001WO-US05674. |
| XX | |
| PR | 17-FEB-2000; 2000US-0183188. |
| XX | |
| PA | (DIAD-) DIADEXUS INC. |
| XX | |
| P1 | Chen S, Sun Y, Macina RA: |
| XX | |
| DR | WPI: 2001-529917/58. |
| XX | |
| DR | P-PSDB; AMG63976. |
| XX | |
| PT | New lung cancer specific gene for the treatment and diagnosis of lung |
| XX | cancer - |
| PS | Claim 1; Page 116-117; 11ppp; English. |
| XX | |
| CC | The present sequence represents a human lung cancer specific gene |
| CC | (LSG), and encodes a polypeptide designated Lng103. LSGs are useful |
| CC | in the treatment and diagnosis of lung cancer. The treatment of lung |
| CC | cancer comprises the administration of a molecule which down regulates |
| CC | the expression of an LSG. An immune response can be mounted against a |
| CC | target cell expressing an LSG. Identification of potential therapeutic |
| CC | agents for use in imaging and treating lung cancer which comprises |
| CC | screening molecules for an ability to bind to or decrease expression |
| CC | of an LSG relative to LSG in the absence of the agent where the ability |
| CC | of a molecule to bind to the LSG or decrease expression of the LSG is |
| CC | indicative of the molecule being useful in imaging and treating lung |
| CC | cancer. |
| XX | |
| SQ | Sequence 2036 BP; 477 A; 586 C; 528 G; 444 T; 1 other; |
| | |
| Query Match 98.6%; Score 1656.2; DB 22; Length 2036; | |
| Best Local Similarity 99.7%; Pred. No. 0; | |
| Matches 1678; Conservative 2; Mismatches 0; Indels 3; Gaps 2. | |
| OY | 1 ggtgctcaggaataagaatttgacttccacagaccactgcgcggaggag- "gagrggagcg 58 |
| DB | 38 ggtgctcaggaataagaatttgacttccacagaccactgcgcggaggagaggaggagcg 97 |
| OY | 59 gcccgaggacttcagcgtgcccagatctgcatctctgcactgtgctgcctctgacaactg 118 |
| DB | 98 gcccgaggacttcagcgtgcccagatctgcatctctgcactgtgctgcctctgacaactg 157 |
| OY | 119 ggaagaatgccggccggcgtagaccttcacacctctctgtgtgttgtgtgtgagccaacttga 178 |
| DB | 158 ggaagaatgccggccggcgtagaccttcacacctctctgtgtgtgtgtgtgagccaacttga 217 |
| OY | 179 tccaagcacaccctcagtcctccactgcagtcttatctctctgcggccaaaagtcatcaagaana 238 |
| DB | 218 tccaagcacaccctcagtcctccactgcagtcttatctctctgcggccaaaagtcatcaagaana 277 |
| OY | 239 agctgacacaaagagctgtaagagaccaaacycaccagacatctctgacgacagctgcgcgtgc 298 |
| DB | 278 agctgacacaaagagctgtaagagaccaaacycaccagacatctctgacgacagctgcgcgtgc 337 |
| OY | 299 tcagtgccatctggggaaaaagccagccggag- atccccgtgtgtgtggcagcctgtggaaca 357 |
| DB | 338 tcagtgccatctggggaaaaagccagccggagagatccccgtgtgtgtggcagcctgtggaaca 397 |
| OY | 358 ccgtctctaagaacatatctgtcgtgaaggtcatcacagctaacatctctccagctgcagag 417 |
| DB | 398 ccgtctctaagaacatatctgtcgtgaaggtcatcacagctaacatctctccagctgcagag 457 |
| OY | 418 tgaagcccttcggccaatgaccagagcgtgctagtaagaatcccccttggaatgtgtgctg 477 |
| DB | 458 tgaagcccttcggccaatgaccagagcgtgctgtaagaatcccccttggaatgtgtgctg 517 |

QY 478 gattcaacaagccctgtgtcaagaacatcgttgagttccacatgacgtgagcccaag 537
|||||
Db 518 gatcaaacgccctctgtgtcaagaacacatcgttgagttccacatgacgtgagcccaag 577
QY 538 ccaacatccgcatgagcaacacgtgcaagtggcccccccgctgtgtccatgtgactgtg 597
|||||
Db 578 ccaccatccgcatgagcaacacgtgcaagtggcccccccgctgtgtccatgtgactgtg 637
QY 598 ccaccagcatgaggagccctggcgcattccaaactgtgtcataagctctccctctgtgaaag 657
|||||
Db 638 ccaccagcatgaggagccctggcgcattccaaactgtgtcataagctctccctctgtgaaag 697
QY 658 ccttagtcaagcagggtcattgaaacctctcgtacccatccctggccaatctagtgaanaac 717
|||||
Db 698 ccttagtcaagcagggtcattgaaacctctcgtacccatccctggccaatctagtgaanaac 757
QY 718 agctgtgtcccggtgatacgagcttccctccatgtgcatgatgcaagacctctgtcagctgtg 777
|||||
Db 758 agctgtgtcccggtgatacgagcttccctccatgtgcatgatgcaagacctctgtcagctgtg 817
QY 778 tgaaggtgccccattccctcagcatctgacacgtctggaagtttgaccttctgtatccctgtcca 837
|||||
Db 818 tgaaggtgccccattccctcagcatctgacacgtctggaagtttgaccttctgtatccctgtcca 877
QY 838 tcaaggtgtgacaacattcaagctctacactgggggccaagtgtgtgactcacagggaag 897
|||||
Db 878 tcaaggtgtgacaacattcaagctctacactgggggccaagtgtgtgactcacagggaag 937
QY 898 tgaaccaagtgttccatcaactctgacagcttccctgacacattcccaacctgtgacaacatcc 957
|||||
Db 938 tgaaccaagtgttccatcaactctgacagcttccctgacacattcccaacctgtgacaacatcc 997
QY 958 cgttcacgctcatcgttgagtcaggaagcgtgtgaaagctgtcagctgtgtgtctctc 1017
|||||
Db 998 cgttcacgctcatcgttgagtcaggaagcgtgtgaaagctgtcagctgtgtgtctctc 1057
QY 1018 cagaagaattcatgttctgttgacactctgtgcttccctgagagtgccccatccgtgtgaag 1077
|||||
Db 1058 cagaagaattcatgttctgttgacactctgtgcttccctgagagtgccccatccgtgtgaag 1117
QY 1078 caagcatcgggtgatcaatgaaagcctgcagataagctgtgatacccaagtcgtga 1137
|||||
Db 1118 caagcatcgggtgatcaatgaaagcctgcagataagctgtgatacccaagtcgtga 1177
QY 1138 agatccctaactcagacacactcccgagtttttaagaccgaagcagatgcaggtgtgccc 1197
|||||
Db 1178 agatccctaactcagacacactcccgagtttttaagaccgaagcagatgcaggtgtgccc 1237
QY 1198 aactgactgtgtgtgaaagtgttccctccagtgaaagccctccgacctgtgttcaacctgtg 1257
|||||
Db 1238 aactgactgtgtgtgaaagtgttccctccagtgaaagccctccgacctgtgttcaacctgtg 1297
QY 1258 gcatcgaagcagctcggaaagctcagttttacaaccaaaagtgtgaccaactatatacctaact 1317
|||||
Db 1298 gcatcgaagcagctcggaaagctcagttttacaaccaaaagtgtgaccaactatatacctaact 1357
QY 1318 tgaataacatcagatctcgatccgagatccagtgatgtaactctgtgattgtgttccaaac 1377
|||||
Db 1358 tgaataacatcagatctcgatccgagatccagtgatgtaactctgtgattgtgttccaaac 1417
QY 1378 ctgagtgtctgtaaaaaacatcatcactgtagatactcaactcactcgtgtgcgaacacaga 1437
|||||
Db 1418 ctgagtgtctgtaaaaaacatcatcactgtagatactcaactcactcgtgtgcgaacacaga 1477
QY 1438 atggcaaatatgaatctgtgggtcccgatgtcatgtgtgaagccttgggattcgaagcag 1497
|||||
Db 1478 atggcaaatatgaatctgtgggtcccgatgtcatgtgtgaagccttgggattcgaagcag 1537
QY 1498 ctgaagcctcactgaccgaaggtatgacctgtgttactccacgctcctctgttgaaccca 1557
|||||
Db 1538 ctgaagcctcactgaccgaaggtatgacctgtgttactccacgctcctctgttgaaccca 1597

QY 1558 gctctcctgtctccacagtgaaagacttgatgtgcagccatcagggaagcgtgtgccagc 1617
|||||
Db 1598 gctctcctgtctccacagtgaaagacttgatgtgcagccatcagggaagcgtgtgtccagc 1657
QY 1618 tggagatattgtgtgtgagctctatagaccatccctctctgtcatcaatataaacacttgcct 1677
|||||
Db 1658 tggagatattgtgtgtgagctctatagaccatccctctctgtcatcaatataaacacttgcct 1717
QY 1678 gtg 1680
|||
Db 1718 gtg 1720

RESULT 5
AAA37057
ID AAA37057 standard; CDNA; 1636 BP.
XX
AC AAA37057;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1357 (UN0706) cDNA sequence SEQ ID NO:127.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PE 01-SEP-1999; 99WO-US20111.
XX
PF 01-SEP-1998; 98US-0098716.
XX
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.

| | | |
|----|--------------|--------------|
| PR | 23-SEP-1998; | 98US-0101377 |
| PR | 23-SEP-1998; | 98US-0101472 |
| PR | 23-SEP-1998; | 98US-0101472 |
| PR | 23-SEP-1998; | 98US-0101475 |
| PR | 23-SEP-1998; | 98US-0101474 |
| PR | 23-SEP-1998; | 98US-0101476 |
| PR | 23-SEP-1998; | 98US-0101477 |
| PR | 23-SEP-1998; | 98US-0101479 |
| PR | 24-SEP-1998; | 98US-0101738 |
| PR | 24-SEP-1998; | 98US-0101741 |
| PR | 24-SEP-1998; | 98US-0101743 |
| PR | 24-SEP-1998; | 98US-0101915 |
| PR | 24-SEP-1998; | 98US-0101916 |
| PR | 29-SEP-1998; | 98US-0102331 |
| PR | 30-SEP-1998; | 98US-0102484 |
| PR | 30-SEP-1998; | 98US-0102487 |
| PR | 30-SEP-1998; | 98US-0102570 |
| PR | 30-SEP-1998; | 98US-0102571 |
| PR | 01-OCT-1998; | 98US-0102684 |
| PR | 01-OCT-1998; | 98US-0102687 |
| PR | 02-OCT-1998; | 98US-0102965 |
| PR | 06-OCT-1998; | 98US-0103358 |
| PR | 06-OCT-1998; | 98US-0103459 |
| PR | 07-OCT-1998; | 98US-0103315 |
| PR | 07-OCT-1998; | 98US-0103328 |
| PR | 07-OCT-1998; | 98US-0103395 |
| PR | 07-OCT-1998; | 98US-0103396 |
| PR | 08-OCT-1998; | 98US-0103401 |
| PR | 08-OCT-1998; | 98US-0103633 |
| PR | 08-OCT-1998; | 98US-0103678 |
| PR | 08-OCT-1998; | 98US-0103679 |
| PR | 14-OCT-1998; | 98US-0104577 |
| PR | 20-OCT-1998; | 98US-0105000 |
| PR | 20-OCT-1998; | 98US-0105002 |
| PR | 21-OCT-1998; | 98US-0105104 |
| PR | 22-OCT-1998; | 98US-0105169 |
| PR | 22-OCT-1998; | 98US-0105266 |
| PR | 26-OCT-1998; | 98US-0105693 |
| PR | 26-OCT-1998; | 98US-0105694 |
| PR | 27-OCT-1998; | 98US-0105807 |
| PR | 27-OCT-1998; | 98US-0105881 |
| PR | 27-OCT-1998; | 98US-0105882 |
| PR | 27-OCT-1998; | 98US-0106062 |
| PR | 28-OCT-1998; | 98US-0106023 |
| PR | 28-OCT-1998; | 98US-0106029 |
| PR | 28-OCT-1998; | 98US-0106032 |
| PR | 28-OCT-1998; | 98US-0106032 |
| PR | 28-OCT-1998; | 98US-0106033 |
| PR | 28-OCT-1998; | 98US-0106033 |
| PR | 29-OCT-1998; | 98US-0106578 |
| PR | 29-OCT-1998; | 98US-0106584 |
| PR | 29-OCT-1998; | 98US-0106580 |
| PR | 30-OCT-1998; | 98US-0106464 |
| PR | 30-OCT-1998; | 98US-0106565 |
| PR | 30-NOV-1998; | 98US-0106502 |
| PR | 30-NOV-1998; | 98US-0106502 |
| PR | 03-NOV-1998; | 98US-0106519 |
| PR | 03-NOV-1998; | 98US-0106532 |
| PR | 03-NOV-1998; | 98US-0106534 |
| PR | 10-NOV-1998; | 98US-0107783 |
| PR | 17-NOV-1998; | 98US-0108775 |
| PR | 17-NOV-1998; | 98US-0108779 |
| PR | 17-NOV-1998; | 98US-0108787 |
| PR | 17-NOV-1998; | 98US-0108788 |
| PR | 17-NOV-1998; | 98US-0108801 |
| PR | 17-NOV-1998; | 98US-0108802 |

PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108844.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108856.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.

(GETH) GENENTECH INC.

PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

DR WPI; 2000-237871/20.
DR P-PSDB; AAY99375.

PT New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions

PS Claim 2; Fig 71; 773pp; English.

CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AA99340 to AA99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding them have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.

SQ Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other;

| | | | | |
|----------------------------|--------|--------------|--------|--------------|
| Query Match | 95.28; | Score 1617; | DB 21; | Length 1636; |
| Best Local Similarity | 99.88; | Pred. NO. 0; | | |
| Matches 1628; Conservative | 1; | Mismatches | 1; | Gaps 1; |

| | | | |
|----|-----|---|-----|
| QY | 51 | grygagcggcgcgagacatccagcgvgccaggtctvgcatctcgacatctgctgcctct | 110 |
| Db | 1 | gagagcgcggcgcgagacatccagcgvgccaggtctvgcatctcgacatctgctgcctct | 60 |
| QY | 111 | gacacctggagaaatgycgcgcgcgcgtgaaacttaacctctctctgtgttgcctggagc | 170 |
| Db | 61 | gacacctggagaaatgycgcgcgcgcgtgaaacttaacctctctctgtgttgcctggagc | 120 |
| QY | 171 | cacctgtatcccaagccacccctcagctcccaactgcaatttcaatctctcggcccaaaatgcat | 230 |
| Db | 121 | cacctgtatcccaagccacccctcagctcccaactgcaatttcaatctctcggcccaaaatgcat | 180 |
| QY | 231 | caaaagaaaagctgacacagagagctggaaggaaccaaaagccacagcatctctgcagcagct | 290 |
| Db | 181 | caaaagaaaagctgacacagagagctggaaggaaccaaaagccacagcatctctgcagcagct | 240 |
| QY | 291 | gcccctgtctcagctgcacatgycgcgggaaaagccagccgcggagagc -atcccctgtctgcggacgct | 348 |
| Db | 241 | gcccctgtctcagctgcacatgycgcgggaaaagccagccgcggagagcatcccctgtctgcggagcct | 300 |
| QY | 350 | ggctgaacacgctccctgaagacatcatctgtgctgaaggtcatcaacagctaaatctctca | 408 |
| Db | 301 | ggctgaacacgctccctgaagacatcatctgtgctgaaggtcatcaacagctaaatctctca | 360 |
| QY | 410 | gctgcaggttggaagccctctgcgaattgacacagagacatctctgtacaggttccccctggagcat | 468 |
| Db | 361 | gctgcaggttggaagccctctgcgaattgacacagagacatctctgtacaggttccccctggagcat | 420 |
| QY | 470 | ggtgcctgagatccaacacgccccctggtccaagacacatctgtgaattccaatgaacgactga | 528 |
| Db | 421 | ggtgcctgagatccaacacgccccctggtccaagacacatctgtgaattccaatgaacgactga | 480 |

QY 530 ggcccaagccacatcccgatcgacaccagtggaagtggcccaaccgcctgctctcag 589
|||||
Db 481 ggcccaagccacatcccgatcgacaccagtggaagtggcccaaccgcctgctctcag 540
QY 590 tgactgtgcacccgcatgggaagctgagcatccaactgctgtcataagctctctct 649
|||||
Db 541 tgactgtgcacccgcatgggaagctgagcatccaactgctgtcataagctctctctct 600
QY 650 ggtgaacgcttaagctaaagctgaagctgaacatctctagtgtccatccctggcccaatctagt 709
|||||
Db 601 ggtgaacgcttaagctaaagctgaagctgaacatctctagtgtccatccctggcccaatctagt 660
QY 710 gaaaacacagctgtgtcccgtagtcgaaggtctctccaatgcatgtatgacacatctct 769
|||||
Db 661 gaaaacacagctgtgtcccgtagtcgaaggtctctccaatgcatgtatgacacatctctct 720
QY 770 gcagctgtgaaggtgcccatttccctcagatctgacccgtctgaggttgaacctctgtta 829
|||||
Db 721 gcagctgtgaaggtgcccatttccctcagatctgacccgtctgaggttgaacctctgtta 780
QY 830 tccctgcatacaagggtgacacatctcactctgaggggccaagttgtttgagctaca 889
|||||
Db 781 tccctgcatacaagggtgacacatctcactctgaggggccaagttgtttgagctaca 840
QY 890 gggaagaagtgaaccaagtgttcaataactctgacagcttccctgacaaatgcccacatctga 949
|||||
Db 841 gggaagaagtgaaccaagtgttcaataactctgacagcttccctgacaaatgcccacatctga 900
QY 950 caacatcccgcttcagctcactcgtgagtgagtgagcagtggaagctgagcagtgctgtgt 1009
|||||
Db 901 caacatcccgcttcagctcactcgtgagtgagtgagcagtggaagctgagcagtgctgtgt 960
QY 1010 gctctctcagaagaattcattggttctgtgtgagctctgtctctcctgaaggtgcccacatg 1069
|||||
Db 961 gctctctcagaagaattcattggttctgtgtgagctctgtctctcctgaaggtgcccacatg 1020
QY 1070 gctgaagtgaagcatcggtggttgcataatgaagaagctgcagataagctggagctcaccca 1129
|||||
Db 1021 gctgaagtgaagcatcggtggttgcataatgaagaagctgcagataagctggagctcaccca 1080
QY 1130 gatcgtgaagatctcactcagacacatcccgagtttltatagacaaaggcatccaca 1189
|||||
Db 1081 gatcgtgaagatctcactcagacacatcccgagtttltatagacaaaggcatccaca 1140
QY 1190 ggtggtcccaactgactcgtctggaagtgttccctcagtgaaagccctccgccttctgtt 1249
|||||
Db 1141 ggtggtcccaactgactcgtctggaagtgttccctcagtgaaagccctccgccttctgtt 1200
QY 1250 caccctgggcatcgaaagcagctcggaaagctcaagttttacaccaaaggtygacaaattat 1309
|||||
Db 1201 caccctgggcatcgaaagcagctcggaaagctcaagttttacaccaaaggtygacaaattat 1260
QY 1310 actcaacttgaataacatcagctctgatcgatccagctgataagctctgggattggctg 1369
|||||
Db 1261 actcaacttgaataacatcagctctgatcgatccagctgataagctctgggattggctg 1320
QY 1370 gttccaactgtatgttctgaaaaaatacatcactagatcatccatccatccctgtgctc 1429
|||||
Db 1321 gttccaactgtatgttctgaaaaaatacatcactagatcatccatccatccctgtgctc 1380
QY 1430 gaaccagaatgacaatgaatgactcggggtcccaagtgtcatatggggaaggtcttggatt 1489
|||||
Db 1381 gaaccagaatgacaatgaatgactcggggtcccaagtgtcatatggggaaggtcttggatt 1440
QY 1490 cgaagcagctgagctcctcaactgacaaagatgaccttgtgcttaactcagcctctctgtg 1549
|||||
Db 1441 cgaagcagctgagctcctcaactgacaaagatgaccttgtgcttaactcagcctctctgtg 1500
QY 1550 gaaacccagctctcctgtctcccaatgaagattgtagtgagcagcatcaagggaaggtctg 1609
|||||
Db 1501 gaaacccagctctcctgtctcccaatgaagattgtagtgagcagcatcaagggaaggtctg 1560

QY 1610 gtcccaagctggaggtatgggtgtgagctctatagacatccctctctgcaataaataaac 1669
|||||
Db 1561 gtcccaagctggaggtatgggtgtgagctctatagacatccctctctgcaataaataaac 1620
QY 1670 acttgctgtg 1680
|||||
Db 1621 acttgctgtg 1631
RESULT 6
AAS46064
ID AAS46064 standard; cDNA; 1636 BP.
XX
AC AAS46064;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human DNA encoding PRO polypeptide sequence #140.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168848-A2.
XX
PD 20-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX

PA (GETH) GENENTECH INC.

PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;

DR WPI: 2001-602746/68.

DR P-PSDB: AA029163.

PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -

PS Claim 2: Fig 279; 774pp: English.

CC Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC primers for PRO polypeptides of the invention. The sequences of the
CC invention can be used to detect the presence of a tumour in a mammal by
CC comparing the level of expression of a PRO polypeptide in a test sample
CC of cells from the animal and a control sample of normal cells, whereby a
CC higher level of expression in the test sample indicates the presence of a
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC pigs, goats and rabbits but are preferably human. The polypeptides can be
CC used to stimulate tumour necrosis factor (TNF) alpha release from human
CC blood, when contacted with it. A specific polypeptide can be used to
CC stimulate the proliferation or differentiation of chondrocyte cells. The
CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.

SO Sequence 1636 BP: 374 A: 492 C: 411 G: 359 T: 0 other:

Query Match 96.2%: Score 1617; DB 22; Length 1636;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1628; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 51 gggagagcgcgagagagacccagctgcccaggtcgtgcatctctgtgcttct 110
DB 1 gaggagcgcgagagagacccagctgcccaggtcgtgcatctctgtgcttct 60
QY 111 gacacctgggaagatgagcgccgtggaaccttacccttctgtggttctgagc 170
DB 61 gacacctgggaagatgagcgccgtggaaccttacccttctgtggttctgagc 120
QY 171 cacccttgatccagacccctcagttccacatgcttcacctctcgcccaaatcat 230
DB 121 cacccttgatccagacccctcagttccacatgcttcacctctcgcccaaatcat 180
QY 231 caaagaagaagctgacacagagctgaagacacaaagccacacagatctctgagagct 290
DB 181 caaagaagaagctgacacagagctgaagacacaaagccacacagatctctgagagct 240
QY 291 gccgctgctcagtgccatcggaagaaagccagccgagag-attccctgtgagcagct 349
DB 241 gccgctgctcagtgccatcggaagaaagccagccgagagcattccctgtgagcagct 300
QY 350 ggtgaacacacgtctcgaagacatcatctgtgctgaaggtcatcaacagtaacctca 409
DB 301 ggtgaacacacgtctcgaagacatcatctgtgctgaaggtcatcaacagtaacctca 360
QY 410 gctgagagtggaagcctcgccaaatgacagagagctgtcagtcacagattcccttgagat 469
DB 361 gctgagagtggaagcctcgccaaatgacagagagctgtcagtcacagattcccttgagat 420
QY 470 ggtgctgattcaaacacccctgtgcaagacatcgtggaattccacatgagacatga 529
DB 421 ggtgctgattcaaacacccctgtgcaagacatcgtggaattccacatgagacatga 480
QY 530 ggcaccaagccacatcgcgaatgacacagatgcaagtggcccaaccgctgtgtctcag 589
DB 481 ggcaccaagccacatcgcgaatgacacagatgcaagtggcccaaccgctgtgtctcag 540

QY 590 tgactgtgccaccagccatggagagcctgacatccaactgctgcaataagctctctcct 649
DB 541 tgactgtgccaccagccatggagagcctgacatccaactgctgcaataagctctctcct 600
QY 650 ggtgaacgccttaagcaacagatcagtaacctcagtgccatccctcccaatcag 709
DB 601 ggtgaacgccttaagcaacagatcagtaacctcagtgccatccctcccaatcag 660
QY 710 gaaaaacagctgtgtcccgatcgagagcttcccttaatggaatgcatgtagaacctct 769
DB 661 gaaaaacagctgtgtcccgatcgagagcttcccttaatggaatgcatgtagaacctct 720
QY 770 gcaagctggtgaaggtggtccatttccctcagatctgacgctctggaattgacctctgta 829
DB 721 gcaagctggtgaaggtggtccatttccctcagatctgacgctctggaattgacctctgta 780
QY 830 tctgtccatcaaggggtgacacatcagctacacccgagggccaagtgtgtgactcaaca 889
DB 781 tctgtccatcaaggggtgacacatcagctacacccgagggccaagtgtgtgactcaaca 840
QY 890 gggaaaggtgacccaagtgttcaataactctgacgcttccctgacaaatgccaccctgga 949
DB 841 gggaaaggtgacccaagtgttcaataactctgacgcttccctgacaaatgccaccctgga 900
QY 950 caacatcccgcttcagctcatctgtagtagagagcgtggaagcgtgagctgtgtgt 1009
DB 901 caacatcccgcttcagctcatctgtagtagagagcgtggaagcgtgagctgtgtgt 960
QY 1010 gctctctcagaagaatcagtgctctgtgtgacgtctgctctcgtgaagatgcccacatg 1069
DB 961 gctctctcagaagaatcagtgctctgtgtgacgtctgctctcgtgaagatgcccacatg 1020
QY 1070 gctgaagtcgaagcatcgtggtgatcaatgaaagcgtcagaaatgagctgagatcaacca 1129
DB 1021 gctgaagtcgaagcatcgtggtgatcaatgaaagcgtcagaaatgagctgagatcaacca 1080
QY 1130 gctcgtggaagatccttaaccagacaccccgagtttcttataagaccagaagggccatgcca 1189
DB 1081 gctcgtggaagatccttaaccagacaccccgagtttcttataagaccagaagggccatgcca 1140
QY 1190 ggtggtcccaactgatactgtctggaagtggttccctcagtgaaagccctccgcttgtt 1249
DB 1141 ggtggtcccaactgatactgtctggaagtggttccctcagtgaaagccctccgcttgtt 1200
QY 1250 caccctgggcatcgaagccagctcgaagctcaagtttcaaccaaaaggtgacccaacttat 1309
DB 1201 caccctgggcatcgaagccagctcgaagctcaagtttcaaccaaaaggtgacccaacttat 1260
QY 1310 actcaactggaatgaatcagctctgtagtgcagcgtgagatgaactctgggattggctg 1369
DB 1261 actcaactggaatgaatcagctctgtagtgcagcgtgagatgaactctgggattggctg 1320
QY 1370 gttccaaactgtatgttctgaaaaaacatcatcactgtagatcatccatccatccctgtgccc 1429
DB 1321 gttccaaactgtatgttctgaaaaaacatcatcactgtagatcatccatccatccctgtgccc 1380
QY 1430 gaacccaagtgtgcaaatgaatcagctcgtggttcccaagtgtcatgttggaagccttggagatt 1489
DB 1381 gaacccaagtgtgcaaatgaatcagctcgtggttcccaagtgtcatgttggaagccttggagatt 1440
QY 1490 cgaagagctgtggttccctcagtcgaacagagatgcccctgtgttaactccagctcctgtgtg 1549
DB 1441 cgaagagctgtggttccctcagtcgaacagagatgcccctgtgttaactccagctcctgtgtg 1500
QY 1550 gaaacccaagctcctcgttcccaagtgaaagactgtgatatgacgccaatcgaaggaagcgtg 1609
DB 1501 gaaacccaagctcctcgttcccaagtgaaagactgtgatatgacgccaatcgaaggaagcgtg 1560
QY 1610 gtcccaagctggagatgaggtgtgagctctatagaccatccctctcgtcaatcaataaac 1669
DB 1561 gtcccaagctggagatgaggtgtgagctctatagaccatccctctcgtcaatcaataaac 1620

OY 1190 ggtggcccaactgacgtgctgtgaagtggttccctccagtgagccctccgcttgtt 1249
|||||
DB 1141 ggtggcccaactgacgtgctgtgaagtggttccctccagtgagccctccgcttgtt 1200
OY 1250 caccctgggcatcgagagccagctcggaagctcagtttaacaccaaagtgaccaattat 1309
|||||
DB 1201 caccctgggcatcgagagccagctcggaagctcagtttaacaccaaagtgaccaattat 1260
OY 1310 actcaactggaataacatcagctgacgtgagtcagcagtgagaaactctggatggctg 1369
|||||
DB 1261 actcaactggaataacatcagctgacgtgagtcagcagtgagaaactctggatggctg 1320
OY 1370 gtccaacctgagtgtctgtgaataacatcactgagatcactccactcctctgtgcc 1429
|||||
DB 1321 gtccaacctgagtgtctgtgaataacatcactgagatcactccactcctctgtgcc 1380
OY 1430 gaaccagaatgagcaaatgaatgaatctgggtgccagtgatcattgtgtgaagccttggatt 1489
|||||
DB 1381 gaaccagaatgagcaaatgaatgaatctgggtgccagtgatcattgtgtgaagccttggatt 1440
OY 1490 cgaagcagctgagctcactgacagcaagatgccccttctgtactcagcctctgtg 1549
|||||
DB 1441 cgaagcagctgagctcactgacagcaagatgccccttctgtactcagcctctgtg 1500
OY 1550 gaaacccagctcctctgtctcccaq tgaagacttgatggcagccatcagggaagctgg 1609
|||||
DB 1501 gaaacccagctcctctgtctcccaq tgaagacttgatggcagccatcagggaagctgg 1560
OY 1610 gtccagctggagtgatgggtgtgagctctatagaccatccctctctgcaatcaataaac 1669
|||||
DB 1561 gtccagctggagtgatgggtgtgagctctatagaccatccctctctgcaatcaataaac 1620
OY 1670 acttgccctgtg 1680
|||||
DB 1621 acttgccctgtg 1631

RESULT 8
AAF54293
ID AAF54293 standard; DNA; 1636 BP.
XX
AC AAF54293;
XX
DT 02-APR-2001 (first entry)
XX
DE DNA encoding protein of the invention #36.
XX
KW Secreted; transmembrane; gene therapy; ss.
XX
OS Unidentified.
XX
PN WO200078961-A1.
XX
PD 28-DEC-2000.
XX
PF 18-FEB-2000; 2000WO-US04342.
XX
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
XX
XX (GENE) GENENTECH INC.
XX
PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

PI Watanabe CK, Williams PM, Wood WI;
XX WPI: 2001-071395/08.
DR
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy -
XX
XX
PS Claim 2; Fig 71; 787pp; English.
XX
CC The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.
XX
SQ Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other;

Query Match 96.2%; Score 1617; DB 22; Length 1636;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1628; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

OY 51 ggtggagcgggcgagagactcagctgcccaggtctgacatctgacttgcctct 110
|:|||||
DB 1 gaggagcgggcgagagactcagctgcccaggtctgacatctgacttgcctct 60
OY 111 gacacctgggaagatggccgcccgtgagacctcaacctctctgtgttctggcagc 170
|||||
DB 61 gacacctgggaagatggccgcccgtgagacctcaacctctctgtgttctggcagc 120
OY 171 caacttgatccaagccacctcagttcccaactcagtttccatctccgcgcccagaatcat 230
|||||
DB 121 caacttgatccaagccacctcagttcccaactcagtttccatctccgcgcccagaatcat 180
OY 231 caaagaaagctgacacagagagctgaagaccacaacgcccacagatctcagcagct 290
|||||
DB 181 caaagaaagctgacacagagagctgaagaccacaacgcccacagatctcagcagct 240
OY 291 gccgctgctcagtgctcagtcgggaaagaccagccggaggt-alcctctgtgtggcagct 349
|||||
DB 241 gccgctgctcagtgctcagtcgggaaagaccagccggaggtalcctctgtgtggcagct 300
OY 350 ggtgaacacgctccctggaagacatcatctgtctggaagttcatcacaagctaaatctcca 409
|||||
DB 301 ggtgaacacgctccctggaagacatcatctgtctggaagttcatcacaagctaaatctcca 360
OY 410 gctcaggtgaaagccctcgcccaatgacagagctgctcagtcagtcagatcccttgagat 469
|||||
DB 361 gctcaggtgaaagccctcgcccaatgacagagctgctcagtcagtcagatcccttgagat 420
OY 470 ggtgctggtatcaacacgcccctgtgtcaagaccatctgtgagttccacatgaactga 529
|||||
DB 421 ggtgctggtatcaacacgcccctgtgtcaagaccatctgtgagttccacatgaactga 480
OY 530 ggcccaagccacatccgcatggaacacagtgcaagtggcccaccgctgtgctccag 589
|||||
DB 481 ggcccaagccacatccgcatggaacacagtgcaagtggcccaccgctgtgctccag 540
OY 590 tgactgtgcacacagccatggagagctgtgacatccaactgcgtcagtaagtcttctct 649
|||||
DB 541 tgactgtgcacacagccatggagagctgtgacatccaactgcgtcagtaagtcttctct 600
OY 650 ggtgaagccttagttaaagcaggtcatggaacctctagtgtccatccctcccaatcagt 709
|||||
DB 601 ggtgaagccttagttaaagcaggtcatggaacctctagtgtccatccctcccaatcagt 660
OY 710 gaaacacagctgtgtcccgtagatcgaggtcttccatgaatggcatgtatgagacctct 769
|||||
DB 661 gaaacacagctgtgtcccgtagatcgaggtcttccatgaatggcatgtatgagacctct 720

Db 716 gaggctgaagaccacaagccacagcatcctcgagcagctgctgctcaatgcatg 775
Oy 310 cgggaaagccagccggagg -atccctgtgctggcagcctgtgtaaacacgctctgaag 368
Db 776 cgggaaagccagccggaggcatccctgtgctggcagcctgtgtaaacacgctctgaag 835
Oy 369 cacatcatctgtgtgaaggtcatcatcaagctaaacatccctcagctgcaagtggaagccctcg 428
Db 836 cacatcatctgtgtgaaggtcatcatcaagctaaacatccctcagctgcaagtggaagccctcg 895
Oy 429 gccaatgacccaggaagctctgaatgaagatcccccgtgacaatggtgctgtgatcaacag 488
Db 896 gccaatgacccaggaagctctgaatgaagatcccccgtgacaatggtgctgtgatcaacag 955
Oy 489 cccctgtgcaagaagacatctgtgaaggttcacatgaagactgaagagcccaagccacatccgc 548
Db 956 cccctgtgcaagaagacatctgtgaaggttcacatgaagactgaagagcccaagccacatccgc 1015
Oy 549 atgagacacagtgcaagtgagcccccagcctgtgtccctcagtgactgtgccaagccat 608
Db 1016 atgagacacagtgcaagtgagcccccagcctgtgtccctcagtgactgtgccaagccat 1075
Oy 609 gggagacccgtgcatacctcgtcgtcataagctctcctcctgtgtgaagccttagctaa 668
Db 1076 gggagacccgtgcatacctcgtcgtcataagctctcctcctgtgtgaagccttagctaa 1135
Oy 669 caggtcatgaacccctctgaatccatccctcgtgccaatctagtgaaacacagctgtgtcc 728
Db 1136 caggtcatgaacccctctgaatccatccctcgtgccaatctagtgaaacacagctgtgtcc 1195
Oy 729 gtgatacgaagcttccctcaatgcatgtatgcagacccctcctcagactgtgtgaagtgccc 788
Db 1196 gtgatacgaagcttccctcaatgcatgtatgcagacccctcctcagactgtgtgaagtgccc 1255
Oy 789 atttccctcagcatgtaccgctctgtgaagtttgactctcgtatcctgcatacaagggctac 848
Db 1256 atttccctcagcatgtaccgctctgtgaagtttgactctcgtatcctgcatacaagggctac 1315
Oy 849 accatcagcctctaccttgaggggcgaagttgtggaactcaacagggaaagtgccaagtg 908
Db 1316 accatcagcctctaccttgaggggcgaagttgtggaactcaacagggaaagtgccaagtg 1375
Oy 909 ttcaataactctgcagcttccctgacaatgcccacactggaacacatcccgctcagccctc 968
Db 1376 ttcaataactctgcagcttccctgacaatgcccacactggaacacatcccgctcagccctc 1435
Oy 969 atcgtgatacgaagcgtgtgaaagctgcagtgctgctgtgtctctcctcagaagaattc 1028
Db 1436 atcgtgatacgaagcgtgtgaaagctgcagtgctgctgtgtctctcctcagaagaattc 1495
Oy 1029 atgtgctgtgtgactgtgtcttctgtagagtgcccaatcggtgaagtcaagcatcg 1088
Db 1496 atgtgctgtgtgactgtgtcttctgtagagtgcccaatcggtgaagtcaagcatcg 1555
Oy 1089 ctgatacattgaaagcgtgcagataagctggaatctacccaatctgtaagaattc 1148
Db 1556 ctgatacattgaaagcgtgcagataagctggaatctacccaatctgtaagaattc 1615
Oy 1149 caggaactcccgag -tttttatagacaaaggcattgccaagtgtgcccacatgactgt 1207
Db 1616 caggaactcccgagtttttatagacaaaggcattgccaagtgtgcccacatgactgt 1675
Oy 1208 gctggaagtgtttccctcagtggaagccctcgcacctgtgttcaacccctgggagatgaagc 1267
Db 1676 gctggaagtgtttccctcagtggaagccctcgcacctgtgttcaacccctgggagatgaagc 1735
Oy 1268 cagctcggaaagctcagtttlaacacaaaggtgacaaacttatactcaactgtgaataacat 1327
Db 1736 cagctcggaaagctcagtttlaacacaaaggtgacaaacttatactcaactgtgaataacat 1795
Oy 1328 cagctcgtatcagatccagctgtgaactctgtggaattgtgctgtccaaactgtatgtct 1387
Db 1796 cagctcgtatcagatccagctgtgaactctgtggaattgtgctgtccaaactgtatgtct 1855

Oy 1388 gaaaaacatcatcatgagatcatcatccatccctgtctgcccgaacacgaatgtgcaaat 1447
Db 1856 gtaaaacatcatcatgagatcatcatccatccctgtctgcccgaacacgaatgtgcaaat 1915
Oy 1448 aagatctgggtgtccagatgtcatatggtgaagccttggatctgagcagctgaatctc 1507
Db 1916 aagatctgggtgtccagatgtcatatggtgaagccttggatctgagcagctgaatctc 1975
Oy 1508 actgaccaaagatgtccctgtgtcttactccagccctctgtgtggaaccacagctccctgt 1567
Db 1976 actgaccaaagatgtccctgtgtcttactccagccctctgtgtggaaccacagctccctgt 2035
Oy 1568 ctcccaatgaagactgtgagcagccatcaagggaagcctgtgtccagctggaatgt 1627
Db 2036 ctcccaatgaagactgtgagcagccatcaagggaagcctgtgtccagctggaatgt 2095
Oy 1628 ggtgtgagctctatagaccatccct 1652
Db 2096 ggtgtgagctctatagaccatccct 2120

RESULT 10
ABAO1989 standard; cDNA: 1035 BP.
XX
AC ABAO1989;
XX
DT 12-FEB-2002 (first entry)
XX
DE Human NOV6b coding sequence.
XX
KW Human; NOV6b; metabolic disorder; neurodegenerative disorder;
KW immune disorder; haematopoietic disorder; developmental disease; cancer;
KW retinal disease; feeding disorder; vaccine; infection; gene therapy;
KW neurological disorder; psychotic disorder; G-protein coupled receptor;
KW cytoskeletal; antidiabetic; virucide; neuroprotective; nootropic;
KW analgesic; antidepressant; antimigraine; anticonvulsant; neuroleptic;
KW antihistaminic; antiallergic; antiinflammatory; anorectic; antiarthritic;
KW antipsoriatic; antithrombotic; antibacterial; fungicide;
KW osteopathic; protozoacide; antidiabetic; antidiabetic; antidiabetic;
KW antiferility; vulnereary; nephrotoxic; hypertensive; hypotensive;
KW salivary gland protein; chromosome 20; ss.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT 1..79
FT 5'UTR
FT 79..1035
FT CDS
FT /*tag= a
FT /*tag= b
FT /*product= "NOV6b"
FT slg_peptide
FT 79..150
FT /*tag= c
FT mat_peptide
FT 151..1032
FT /*tag= d
FN WO200174851-A2.
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US10039.
PE
XX 30-MAR-2000; 2000US-193205P.
PR 30-MAR-2000; 2000US-193339P.
PR 05-APR-2000; 2000US-195343P.
PR 06-APR-2000; 2000US-195005P.
PR 06-APR-2000; 2000US-195088P.
PR 10-APR-2000; 2000US-195792P.
PR 11-APR-2000; 2000US-196556P.
PR 13-APR-2000; 2000US-197081P.
PR 14-APR-2000; 2000US-197087P.
PR 14-APR-2000; 2000US-197525P.

06-APR-2000; 2000US-195088P.
10-APR-2000; 2000US-195792P.
PR 11-APR-2000; 2000US-196556P.
PR 13-APR-2000; 2000US-197081P.
PR 14-APR-2000; 2000US-197087P.
PR 14-APR-2000; 2000US-197525P.
PR 29-MAR-2001; 2001US-0823187.
XX
XX (CURA-) CURAGEN CORP.
XX
PI Mayumder K., Spaderna SK, Taupier RJ, Padigaru M, Burgess CE;
PI Shlunkes RA, Szytek KA, Liu X, Patturajan M, Gusev VI;
XX
XX WPI: 2001-626379/72.
DR P-PSDB: AAM47214.
XX
XX
PT New G protein-coupled receptor related polypeptides and polynucleotides
PT for diagnosis, prevention and treatment of metabolic,
PT neurodegenerative, retinal, immune, hematopoietic disorders, diabetes,
PT obesity and infections -
XX
XX
PS Claim 8; Page 54; 194pp: English.
XX
XX The present invention provides the protein and coding sequences of novel
CC human G-protein coupled receptors, designated NOV1, NOV2, NOV3, NOV4,
CC NOV5a, NOV5b, NOV6a, NOV6b, NOV7, NOV8a, NOV8b, NOV9 and NOV10. These
CC can be used in the treatment of NOVY related diseases, including cancer,
CC metabolic, neurodegenerative, immune, haematopoietic, developmental,
CC retinal, feeding, neurological and psychiatric diseases and disorders and
CC infections. The present sequence is the NOV6a coding sequence, the gene
CC for which is found on chromosome 20. The NOV6a protein shares homology
CC with salivary gland proteins.
XX
XX Sequence 963 BP; 217 A; 314 C; 237 G; 195 T; 0 other;

| Query Match | 43.1% | Score 724.2; | DB 22; | Length 963; |
|-----------------------|---|---------------------|------------|-------------|
| Best Local Similarity | 90.1%; | Pred. No. 2.7e-178; | | |
| Matches 842; | Conservative 0; | Mismatches 38; | Indels 55; | Gaps 4; |
| QY 124 | atggcgcgcgcgtggaacttaaccccttcctgtgtgtgtgtgctgagcagccacttgatccaa | 183 | | |
| Db 1 | atggcgcgcgcgtggaacttaaccccttcctgtgtgtgtgtgctgagcagccacttgatccaa | 60 | | |
| QY 184 | gccacccctcagtcctcaactgcagttcttcatactccgcgcgcacaaagatcatcaagaataagtcg | 243 | | |
| Db 61 | gccacccctcagttcccaactgcagttcttcatactccgcgcgcacaaagatcatcaagaataagtcg | 120 | | |
| QY 244 | acacagagagctgaagagaccacacacgcgcacacagatctctgcagcagctgcgcgtgtcagtt | 303 | | |
| Db 121 | acacagagagctgaagagaccacacacgcgcacacagatctctgcagcagctgcgcgtgtcagtt | 180 | | |
| QY 304 | gccatcgcggggaagaagccagccgcggagag-atccctgtgtgtgtgagcagcctgtgtgaacacgtc | 362 | | |
| Db 181 | gccatcgcggggaagaagccagccgcggagcattccctgtgtgtgtgagcagcctgtgtgaacacgtc | 240 | | |
| QY 363 | ctggaagcacatca-----tctggtgtgaagttcatcaacagctcaacatctctcagctgcag | 416 | | |
| Db 241 | ctggaagcacatcaaccccatccaggtctgaagttcatcaacagctcaacatctctcagctgcag | 300 | | |
| QY 417 | gtgaagccctcgcgcacatgtacagagagctgtcgtagtccaagaatcccccctgacatgtgtgct | 476 | | |
| Db 301 | gtgaagccctcgcgcacatgtacagagagagctgtcgtagtccaagaatcccccctgacatgtgtgct | 360 | | |
| QY 477 | ggatccaacacgcgcctgtgtccaagaaccatcgtgtgaagttccaatgtacagactgaagcccaa | 536 | | |
| Db 361 | ggatccaacacgcgcctgtgtccaagaaccatcgtgtgaagttccaatgtacagactgaagcccaa | 420 | | |
| QY 537 | gccacacatccgcgatgtgaacacagctgcagaatgtgcccacccgcctgtgtctcaagtgaactgt | 596 | | |
| Db 421 | gccacacatccgcgatgtgaacacagctgcagaatgtgcccacccgcctgtgtctcaagtgaactgt | 480 | | |
| QY 587 | gccacacacgatgaagaagctgaagatccaactgtactgaataaactctccttctgtgtgaagc | 656 | | |

| | | | |
|----|------|--|------|
| Db | 481 | gcaacaacgacacatgagagccttcgcatcccaactcgtctgataagaacctctcttccttcgttgaaac | 540 |
| Oy | 657 | gaccttagtaagaagatgataaactccttaatgtgcatccctgcaccccaattagtgaaaaac | 716 |
| Db | 541 | gccttaagctaaagcaagtgataagaaacctctctagtgcataccctgcgccaattctagtgtaaaaac | 600 |
| Oy | 717 | cagctgtgtcccgatcatcgagctcttccttcaatgtgcatgtatgtagagaacctcttcgacgtg | 776 |
| Db | 601 | cagctgtgtcccgatcatcgagctcttccttcaatgtgcatgtatgtagagaacctcttcgacgtg | 660 |
| Oy | 777 | gtgaaagtgcccatcttcctctcagcatgtgacccgtctgtgaagtgttgacctctgtatccctgcc | 836 |
| Db | 661 | gtgaag-----gtaggtgtctctgtcctctctccacatttctct---700 | 700 |
| Oy | 837 | atcaaagggtgacaccatctcagctctaacctgtggggcgaagttgtgtgaactcaagvgaag | 896 |
| Db | 701 | -----tactacggagtgctgagctctccacgaacccggaaag | 732 |
| Oy | 897 | gtgaccagtggttcaataaactctgcagcttccttccttgacaatgcccacccttgagaaacatc | 956 |
| Db | 733 | gtgaccagtggttcaataaactctgcagcttccttccttgacaatgcccacccttgagaaacatc | 792 |
| Oy | 957 | ccgttcaaccccatctctgtagtcacagacgtgtgtgaagcgtgcagtgctgtcgtctctct | 1016 |
| Db | 793 | ccgttcaaccccatctctgtagtcacagacgtgtgtgaagcgtgcagtgctgtcgtctctct | 852 |
| Oy | 1017 | ccagaagaattcatgtctctgtttgagactgtgtct | 1051 |
| Db | 853 | ccagaagaattcatgtctctgtttgagactgtgtct | 887 |

| | | |
|----------|---|--------------------------|
| RESULT | 12 | |
| AA297114 | | |
| ID | AA297114 | standard; cDNA: 1172 BP. |
| XX | | |
| AC | AA297114; | |
| XX | | |
| DT | 19-APR-2000 | (first entry) |
| DE | Human secreted protein gene 5 cDNA clone HBHMA23, SEQ ID NO:106. | |
| XX | | |
| KW | Human; secreted protein; cancer; tumour; developmental abnormality; | |
| KW | fetal deficiency; blood disorder; immune system disorder; inflammation; | |
| KW | autoimmune disease; allergy; Alzheimer's disease; cognitive disorder; | |
| KW | scleroderma; arthritis; asthma; psoriasis; sepsis; skin disorder; | |
| KW | atherosclerosis; diabetes; cardiovascular disorder; kidney disorder; | |
| KW | digestive disorder; endocrine disorder; infection; AIDS; leukaemia; | |
| KW | therapy; ds. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| PN | W09966041-A1. | |
| XX | | |
| PD | 23-DEC-1999. | |
| XX | | |
| PE | 15-JUN-1999; | 99WO-US13418. |
| XX | | |
| PR | 16-JUN-1998; | 98US-0089507. |
| PR | 16-JUN-1998; | 98US-0089508. |
| PR | 16-JUN-1998; | 98US-0089509. |
| PR | 16-JUN-1998; | 98US-0089510. |
| PR | 22-JUN-1998; | 98US-0090112. |
| PR | 22-JUN-1998; | 98US-0090113. |
| XX | | |
| PA | (HUMA-) | HUMAN GENOME SCI INC. |
| XX | | |
| PI | Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence RA; | |
| PI | Soppet DR, Brewer LA, Endress GA, Carter KC, Mucensti M, Ebner R; | |
| PI | Lalleur DM, Olsen HS, Shi Y, Moore PA, Komatsoulis G; | |
| XX | | |
| WP1 | 2000-106100/09. | |
| DR | P-PSDB; AAY86310. | |

XX New isolated human genes and the secreted polypeptides they encode,
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders -

PS Claim 1: Page 374; 586pp; English.

CC AA97019 to AA97137 represent 94 isolated human secreted protein genes.
 CC AAY86215 to AAY86333 are the secreted proteins encoded by the 94 human
 CC genes. This sequence represents a fragment of one of the human secreted
 CC proteins. The genes and their corresponding secreted polypeptides are
 CC useful for preventing, treating or ameliorating medical conditions,
 CC e.g., by protein or gene therapy. Also pathological conditions can be
 CC diagnosed by determining the presence of mutations in the new genes. Specific
 CC uses are described for each of the 94 genes, based on which tissues they
 CC are most highly expressed in, and include developing products for the
 CC diagnosis or treatment of cancer, tumours, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
 CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
 CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
 CC disorders, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The sequences shown in AAY86334 to AAY86585 represent fragments of the
 CC secreted proteins.

CC Sequence 1172 BP; 287 A; 358 C; 281 G; 238 T; 8 other;

Query Match 38.2%; Score 641; DB 21; Length 1172;

Best Local Similarity 99.1%; Pred. NO. 1.3e-156;
 Matches 641; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 56 gcggcgagagactccagcgtgccaggtctgcacatctgcagctggccctctgcaac 115
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 2 gcggcgagagactccagcgtgccaggtctgcacatctgcagctggccctctgcaac 61
 QY 116 ctgggaagatggtccggccgtggaactcaccctctctgtgttctgctggcagccact 175
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 62 ctgggaagatggtccggccgtggaactcaccctctctgtgttctgctggcagccact 121
 QY 176 tgatccaaagcaccctagttccacttcagttcattcattctctggcccaaatatcaag 235
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 122 tgatccaaagcaccctagttccacttcagttcattcattctctggcccaaatatcaag 181
 QY 236 aaaagctgaacacagagctgaagaaacacaaagccaccagcatcctcagcagctgcgc 295
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 182 aaaagctgaacacagagctgaagaaacacaaagccaccagcatcctcagcagctgcgc 241
 QY 296 tgcctaatgcatcgtgggaaagcagcagcgagagatccctgtgtctggcgagcctgtgaa 355
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 242 tgcctaatgcatcgtgggaaagcagcagcgagagatccctgtgtctggcgagcctgtgaa 301
 QY 356 caccgtcctaaagacatcatctgtgctgaaggtcatcacagctcctaactctcagctgca 415
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 302 caccgtcctaaagacatcatctgtgctgaaggtcatcacagctcctaactctcagctgca 361
 QY 416 ggtgaagcctcggtgcaatgaacagagctgtcagtcagtcacatccctctggacatgtgagc 475
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 362 ggtgaagcctcggtgcaatgaacagagctgtcagtcagtcacatccctctggacatgtgagc 421
 QY 476 tggattcaaacagccctctgttcaagaccatcgtgagattccacatgaagcagtgagccca 535
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 422 tggattcaaacagccctctgttcaagaccatcgtgagattccacatgaagcagtgagccca 481
 QY 536 agccaccatccgcatggaacccagtgcaagtggcccccacccgctgtctcagtgactg 595
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 482 agccaccatccgcatggaacccagtgcaagtggcccccacccgctgtctcagtgactg 541
 QY 596 tgcacacagcatgtgagcctcgtcgaatccaactgtcagatcaagctctcctctctgttga 655
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 542 tgcacacagcatgtgagcctcgtcgaatccaactgtcagatcaagctctcctctctgttga 601

QY 656 cgccttagtaagcaggtcatgaactcctagtgcacctcctgccca 702
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 602 cgccttagtaagcaggtcatgaactcctagtgcacctcctgccca 648

RESULT 13

AAFI8254
 ID AAF18254 standard; DNA; 1177 BP.

AC AAF18254;

DT 14-MAR-2001 (first entry)

DE Lung cancer associated polynucleotide sequence SEQ ID 273.

XX Human; lung cancer associated protein; neuroprotective; cytoskeletal;
 KW cardioactive; immunomodulatory; muscular active; vulnerrary;
 KW gastrointestinal; nephrotropic; antiinfective; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.

OS Homo sapiens.

PN W020005180-A2.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05918.

PR 12-MAR-1999; 9905-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PT (ROSE/) ROSEN C A.

PI Ruben SM;

DR WPI; 2000-587514/55.

DT P-PADB; AAB58378.

PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 PS Claim 1: Page 732; 1425pp; English.

CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytoskeletal; cardioactive;
 CC immunomodulatory; muscular active general; vulnerrary; gastrointestinal
 CC general; nephrotropic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.

SO Sequence 1177 BP; 290 A; 361 C; 287 G; 239 T; 0 other;

Query Match 37.8%; Score 635.8; DB 21; Length 1177;

Best Local Similarity 99.5%; Pred. NO. 2.9e-155;
 Matches 648; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 53 ggaagcgagcagagactccagcgtgccaggtctgcacatcctcagctgtgcctctga 112

|||||
Db 2 ggaagcgccgagagctccagcgctgcgcaagctcctgcatccctgcaactgctgccccctga 61
QY 113 caacttgggaagatgagcgccgctggaacttcaaccttctctgtgttctgctggagcca 172
Db 62 caacttgggaagatgagcgccgctggaacttcaaccttctctgtgttctgctggagcca 121
QY 173 ccttgatccaagccacccctcagctccacatgcttctctcctcctggcccaaaagtatca 232
Db 122 ccttgatccaagccacccctcagctccacatgcttctctcctcctggcccaaaagtatca 181
QY 233 aagaaaagctgacacagagagctgaaagagaccacacagccacacatcctgacagactgc 292
Db 182 aagaaaagctgacacagagagctgaaagagaccacacagccacacatcctgacagactgc 241
QY 293 cgctgtcagtgccatgctgggaaagacagccggagag-atccctgtgtctggcagcctgg 351
Db 242 cgctgtcagtgccatgctgggaaagacagccggagagacatccctgtgtctggcagcctgg 301
QY 352 tgaacacgctcctgaaagacatcatctgtgctgaaagtcatcagctaaatcctcctcagc 411
Db 302 tgaacacgctcctgaaagacatcatctgtgctgaaagtcatcagctaaatcctcctcagc 361
QY 412 tgcaggttgaagccctcgcgcaatgacacagagagctgctagtcacagatccccctggacatgg 471
Db 362 tgcaggttgaagccctcgcgcaatgacacagagagctgctagtcacagatccccctggacatgg 421
QY 472 tgcgtgtgattcaacagccctcgtgtgtaagaacatcgtgtgagttcacaatgacagactgag 531
Db 422 tgcgtgtgattcaacagccctcgtgtgtaagaacatcgtgtgagttcacaatgacagactgag 481
QY 532 cccaagccacatcgcagctggaacacagctgcaagtggcccccacccgctgctcctagctg 591
Db 482 cccaagccacatcgcagctggaacacagctgcaagtggcccccacccgctgctcctagctg 541
QY 592 actgtgccaacagcagctggaagccttgcagatccaactgtctgacataagctctcctctgg 651
Db 542 actgtgccaacagcagctggaagccttgcagatccaactgtctgacataagctctcctctgg 601
QY 652 tgaagccttgaagcagctgacatgaaactcctctagtcacatccctcgccca 702
Db 602 tgaagccttgaagcagctgacatgaaactcctctagtcacatccctcgccca 652
RESULT 14
AAZ97023
ID AAZ97023 standard; cDNA; 1175 BP.
XX
AC AAZ97023;
XX
DT 19-APR-2000 (first entry)
XX
DE Human secreted protein gene 5 cDNA clone HBHMA23, SEQ ID NO:15.
XX
KW Human; secreted protein; cancer; tumour; developmental abnormality;
KW foetal deficiency; blood disorder; immune system disorder; inflammation;
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
KW therapy; ds.
XX
OS Homo sapiens.
XX
PN WO966041-A1.
XX
PD 23-DEC-1999.
XX
PF 15-JUN-1999; 99MO-US13418.
XX
PR 16-JUN-1998; 98US-0089507.
PR 16-JUN-1998; 98US-0089508.
PR 16-JUN-1998; 98US-0089509.

PR 16-JUN-1998; 98US-0089510.
PR 22-JUN-1998; 98US-0090112.
PR 22-JUN-1998; 98US-0090113.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA;
PI Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Edner R;
PI Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsuolis G;
XX
XX WPI: 2000-106100/09.
DR P-PSDB: AAY86219.
XX
PT New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 1: Page 318; 586pp; English.
CC AAZ97019 to AAZ97137 represent 94 isolated human secreted protein genes.
CC AAY86215 to AAY86333 are the secreted proteins encoded by the 94 human
CC genes. This sequence represents a fragment of one of the human secreted
CC proteins. The genes and their corresponding secreted polypeptides are
CC useful for preventing, treating or ameliorating medical conditions,
CC e.g., by protein or gene therapy. Also pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new genes. Specific
CC uses are described for each of the 94 genes, based on which tissues they
CC are most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, tumours, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC disorders, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The sequences shown in AAY86334 to AAY86585 represent fragments of the
XX secreted proteins.
SQ Sequence 1175 BP; 290 A; 359 C; 286 G; 239 T; 1 other:
Query Match 37.8%; Score 634.8; DB 21; Length 1175;
Best local Similarity 99.5%; Pred. No. 5; 2e-155;
Matches 647; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 54 gagcgcgagagagctccagcgctgcccaggtctgcatcctgcaactgtgcctctgac 113
Db 1 gagcgcgagagagctccagcgctgcccaggtctgcatcctgcaactgtgcctctgac 60
QY 114 acctgggaagatggccgcccgtggacatcccttctctgtgttctgtgcaagccac 173
Db 61 acctgggaagatggccgcccgtggacatcccttctctgtgttctgtgcaagccac 120
QY 174 ctgacccaagcaccctcagctccacatgcttccatcctctgcgcccacaaatcaca 233
Db 121 ctgacccaagcaccctcagctccacatgcttccatcctctgcgcccacaaatcaca 180
QY 234 agaaaagctgacacagagctgaaagacacacagccacacagatcctctgacagctgac 293
Db 181 agaaaagctgacacagagctgaaagacacacagccacacagatcctctgacagctgac 240
QY 294 gctgtcagtgccatgctgggaaagacagccggagag-atccctgtgtctggagcctggt 352
Db 241 gctgtcagtgccatgctgggaaagacagccggagagatccctgtgtctggagcctggt 300
QY 353 gaacacgctccttgaacacacatcctgctgtgaagtcatacagacacacatcctcagct 412
Db 301 gaacacgctccttgaacacacatcctgctgtgaagtcatacagacacacatcctcagct 360
QY 413 gcaggtgaagccctcgcccaatgacacagagctgtctagtcagaatcccccctgacatggt 472
Db 361 gcaggtgaagccctcgcccaatgacacagagctgtctagtcagaatcccccctgacatggt 420

QY 473 ggctgattcaacagccctgtgtaagaccatcgtgagttccacatgaagcagc 532
|||||
Db 421 ggctgattcaacagccctgtgtaagaccatcgtgagttccacatgaagcagc 480
|||||
QY 533 ccaagccacacatccgcatggaacacagtcgaagtcgccccacccgctgtctcctcagta 592
|||||
Db 481 ccaagccacacatccgcatggaacacagtcgaagtcgccccacccgctgtctcctcagta 540
|||||
QY 593 ctgtgcccacacagcattggaagcctgcatcccaactgtcgtcgtcctcctcctcctcgt 652
|||||
Db 541 ctgtgcccacacagcattggaagcctgcatcccaactgtcgtcgtcctcctcctcctcgt 600
|||||
QY 653 gaagccttaagcctgaagcagtcataaactcctcctcctcctcctcctcctcctcctc 702
|||||
Db 601 gaagccttaagcctgaagcagtcataaactcctcctcctcctcctcctcctcctcctc 650
|||||
RESULT 15
AAH30554
ID AAH30554 standard; cDNA; 420 BP.
XX
AC AAH30554;
XX
DT 27-JUL-2001 (first entry)
XX
DE Human colon cancer cell line Km12L4-A cDNA library derived sequence #488.
XX
KW Human; diagnosis: colon cancer; cancer: malignant; chromosome mapping:
KM detection; colon cancer cell line Km12L4-A, ss.
XX
OS Homo sapiens.
XX
FN W0200018916-A2.
XX
PD 06-APR-2000.
XX
PF 23-SEP-1999; 99WO-US22226.
XX
PR 28-SEP-1998; 98US-0102161.
PR 28-SEP-1998; 98US-0102180.
PR 29-SEP-1998; 98US-0102380.
PR 08-OCT-1998; 98US-0103815.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PI (HYSE-) HYSEO INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Iamson G, Drmanac R, Ckvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
DR WPI; 2000-293155/25.
XX
PT Polynucleotide library comprising 1079 defined sequences, useful in
PT the form of an array to detect cancer or susceptibility to cancer -
XX
PS Claim 1; Page 323-324; 502pp; English.
XX
XX
CC The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (I);
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides
CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive

CC interventions. Polynucleotide probes based on the disclosed sequences
CC are useful for chromosome mapping and detection of transcription levels.
CC The 1079 polynucleotide sequences were derived from a human colon cancer
CC cell line Km12L4-A cDNA library.

XX Sequence 420 BP; 97 A; 116 C; 108 G; 99 T; 0 other;

Query Match 25.0%; Score 420; DB 21; Length 420;
Best Local Similarity 100.0%; Pred. No. 2.5e-99;

Matches 420; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 870 gccaatgttgatgactcaaggaaggtgacaaagtgttaactcgtcagctcc 929
|||||
Db 1 gccaatgttgatgactcaaggaaggtgacaaagtgttaactcgtcagctcc 60
|||||
QY 930 ctgacaatgcccacccctggaacacatccglttcagctcactcgtgagtcagcagctgtg 989
|||||
Db 61 ctgacaatgcccacccctggaacacatccglttcagctcactcgtgagtcagcagctgtg 120
|||||
QY 990 aaagctgcaagt 1049
|||||
Db 121 aaagctgcaagt 180
|||||
QY 1050 ctctcctgagagtgcccatcgctggaagtcgaagtcggctggtatgaatgaagaagctgca 1109
|||||
Db 181 ctctcctgagagtgcccatcgctggaagtcgaagtcggctggtatgaatgaagaagctgca 240
|||||
QY 1110 gataagctggatctaccagatcgtggaagatcctaactcagacactccgagttttc 1169
|||||
Db 241 gataagctggatctaccagatcgtggaagatcctaactcagacactccgagttttc 300
|||||
QY 1170 atagaccaagcagcattgccaagtggtgcccactatcgtgtgtgaaagtgttccctcagt 1229
|||||
Db 301 atagaccaagcagcattgccaagtggtgcccactatcgtgtgtgaaagtgttccctcagt 360
|||||
QY 1230 gaagccctccgccccttgttaccacctggtgagatgaagccagcgcggaagctcagtttac 1289
|||||
Db 361 gaagccctccgccccttgttaccacctggtgagatgaagccagcgcggaagctcagtttac 420
|||||

Search completed: September 16, 2002, 02:55:06
Job time: 7384 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 00:51:17 : Search time 89 Seconds
(without alignments)
4636.677 Million cell updates/sec

Title: US-09-700-770-2
Perfect score: 1680
Sequence: 1 gctgctgagatataagtt.....tcaataaacacttgctctg 1680

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PCCTS.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|---------|-------|-------------------|
| 1 | 41.6 | 2.5 | 44377 | 2 | US-08-804-227C-7 |
| 2 | 41.6 | 2.5 | 44377 | 2 | US-08-804-198-1 |
| 3 | 39.6 | 2.4 | 4245 | 4 | US-09-276-531-16 |
| C | 38.4 | 2.3 | 4403765 | 4 | US-09-103-840A-2 |
| 5 | 38.2 | 2.3 | 1931 | 2 | US-09-130-114-2 |
| 6 | 37.6 | 2.2 | 1301 | 4 | US-08-804-166-7 |
| 7 | 37.6 | 2.2 | 1301 | 4 | US-08-910-991-7 |
| 8 | 37.4 | 2.2 | 3241 | 4 | US-09-434-288-11 |
| 9 | 37.2 | 2.2 | 1241 | 1 | US-08-471-033-39 |
| 10 | 37.2 | 2.2 | 1241 | 1 | US-08-471-033-42 |
| 11 | 37.2 | 2.2 | 1241 | 2 | US-08-471-044-39 |
| 12 | 37.2 | 2.2 | 1241 | 2 | US-08-471-044-42 |
| 13 | 37.2 | 2.2 | 1241 | 2 | US-08-463-483A-39 |
| 14 | 37.2 | 2.2 | 1241 | 2 | US-08-463-483A-42 |
| 15 | 37.2 | 2.2 | 1241 | 2 | US-08-471-046A-39 |
| 16 | 37.2 | 2.2 | 1241 | 2 | US-08-471-046A-42 |
| 17 | 37.2 | 2.2 | 1241 | 2 | US-08-470-566B-39 |
| 18 | 37.2 | 2.2 | 1241 | 2 | US-08-470-566B-42 |
| 19 | 37.2 | 2.2 | 1241 | 2 | US-08-469-334-39 |
| 20 | 37.2 | 2.2 | 1241 | 2 | US-08-469-334-42 |
| 21 | 37.2 | 2.2 | 1241 | 3 | US-09-300-529-39 |
| 22 | 37.2 | 2.2 | 1241 | 3 | US-09-300-529-42 |
| 23 | 37.2 | 2.2 | 1358 | 1 | US-08-471-033-45 |
| 24 | 37.2 | 2.2 | 1358 | 2 | US-08-471-044-45 |
| 25 | 37.2 | 2.2 | 1358 | 2 | US-08-463-483A-45 |
| 26 | 37.2 | 2.2 | 1358 | 2 | US-08-471-046A-45 |
| 27 | 37.2 | 2.2 | 1358 | 2 | US-08-470-566B-45 |

| | | | | | | |
|----|------|-----|------|---|-------------------|-------------------|
| 28 | 37.2 | 2.2 | 1358 | 2 | US-08-469-334-45 | Sequence 45, Appl |
| 29 | 37.2 | 2.2 | 1358 | 3 | US-09-300-529-45 | Sequence 45, Appl |
| 30 | 37.2 | 2.2 | 1389 | 1 | US-08-471-033-27 | Sequence 27, Appl |
| 31 | 37.2 | 2.2 | 1389 | 2 | US-08-471-044-27 | Sequence 27, Appl |
| 32 | 37.2 | 2.2 | 1389 | 2 | US-08-463-483A-27 | Sequence 27, Appl |
| 33 | 37.2 | 2.2 | 1389 | 2 | US-08-471-046A-27 | Sequence 27, Appl |
| 34 | 37.2 | 2.2 | 1389 | 2 | US-08-470-566B-27 | Sequence 27, Appl |
| 35 | 37.2 | 2.2 | 1389 | 2 | US-08-469-334-27 | Sequence 27, Appl |
| 36 | 37.2 | 2.2 | 1389 | 3 | US-09-300-529-27 | Sequence 27, Appl |
| C | 37.2 | 2.2 | 1995 | 1 | US-08-425-063-3 | Sequence 3, Appl |
| 38 | 37.2 | 2.2 | 1995 | 2 | US-08-317-844B-3 | Sequence 3, Appl |
| 39 | 37.2 | 2.2 | 4031 | 1 | US-08-471-033-49 | Sequence 49, Appl |
| 40 | 37.2 | 2.2 | 4031 | 2 | US-08-471-044-49 | Sequence 49, Appl |
| 41 | 37.2 | 2.2 | 4031 | 2 | US-08-463-483A-49 | Sequence 49, Appl |
| 42 | 37.2 | 2.2 | 4031 | 2 | US-08-471-046A-49 | Sequence 49, Appl |
| 43 | 37.2 | 2.2 | 4031 | 2 | US-08-470-566B-49 | Sequence 49, Appl |
| 44 | 37.2 | 2.2 | 4031 | 2 | US-08-469-334-49 | Sequence 49, Appl |
| 45 | 37.2 | 2.2 | 4031 | 3 | US-09-300-529-49 | Sequence 49, Appl |

ALIGNMENTS

RESULT 1
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhlthoss, Stuart A.
; APPLICANT: Rustock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text. only
; CURRENT APPLICATION DATA:
; FILING DATE: February 21, 1997
; APPLICATION NUMBER: US/08/804,227C
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:

NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match 2.5%; Score 41.6; DB 2; Length 44377;
Best Local Similarity 49.5%; Pred. No. 0.5;
Matches 107; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 332 cccctgctggcagcctggtgtaaacgcgtccctgaagacatcatctgctgaagtcagt 391
DB 43661 CCTCGAAGCTGGCGGGCCGGGATCACCGCCGACGACGACGTCGTGTCGTGGCCGCCG 43720

QY 392 cacaagctaacatctcccaagctcaggtgaagccctcgccaatgaccagagctgtagt 451
DB 43721 CGGACGGGAGATCCCGGGCCGGAGCCGCTCGGGGGCGCCGACGACGACGTCGTGTCGTGGCCGCCG 43780

QY 452 caagatcccccctggaatggtgctgattcaacagccctggtcaagaccatcgtgta 511
DB 43781 CGCCCTCCCGCTGATCCGCTGGCCGACACACCGGTCAGGTCCAGGCGCGCGCT 43840

QY 512 gtccacatgaagctgagggcccaagccaccatcgcg 547
DB 43841 GCCCGGCTGGGGCGCGGAGGACCGCCCTGACCCG 43876

RESULT 2
US-08-804-198-1

; Sequence 1, Application US/08804198
; Patent No. 5945320

; GENERAL INFORMATION:

; APPLICANT: Burgett, Stanley G.

; APPLICANT: Kustoss, Stuart A.

; APPLICANT: Rao, Nagaraja R.

; APPLICANT: Richardson, Mark A.

; APPLICANT: Rostock, Paul R., Jr.

; TITLE OF INVENTION: PLATENOILIDE SYNTHASE GENE

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PAUL R. CANTRELL 1138

; STREET: LILLY CORPORATE CENTER

; CITY: INDIANAPOLIS

; STATE: IN

; COUNTRY: USA

; ZIP: 46285

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Macintosh

; OPERATING SYSTEM: Macintosh 7.0

; SOFTWARE: Microsoft Word 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/804,198

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: CANTRELL, PAUL R.

; REGISTRATION NUMBER: 36,470

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-3885

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 44377 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 350..14002

FEATURE:

NAME/KEY: CDS

LOCATION: 14046..20036

FEATURE:

NAME/KEY: CDS

LOCATION: 20110..31284

FEATURE:

NAME/KEY: CDS

LOCATION: 31329..36071

FEATURE:

NAME/KEY: CDS

LOCATION: 36155..41830

US-08-804-198-1

Query Match 2.5%; Score 41.6; DB 2; Length 44377;
Best Local Similarity 49.5%; Pred. No. 0.5;
Matches 107; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 332 cccctgctggcagcctggtgtaaacgcgtccctgaagacatcatctgctgaagtcagt 391
DB 43661 CCTCGAAGCTGGCGGGCCGGGATCACCGCCGACGACGACGTCGTGTCGTGGCCGCCG 43720

QY 392 cacaagctaacatctcccaagctcaggtgaagccctcgccaatgaccagagctgtagt 451
DB 43721 CGGACGGGAGATCCCGGGCCGGAGCCGCTCGGGGGCGCCGACGACGACGTCGTGTCGTGGCCGCCG 43780

QY 452 caagatcccccctggaatggtgctgattcaacagccctggtcaagaccatcgtgta 511
DB 43781 CGCCCTCCCGCTGATCCGCTGGCCGACACACCGGTCAGGTCCAGGCGCGCGCT 43840

QY 512 gtccacatgaagctgagggcccaagccaccatcgcg 547
DB 43841 GCCCGGCTGGGGCGCGGAGGACCGCCCTGACCCG 43876

RESULT 3

US-09-276-531-16

; Sequence 16, Application US/09276531

; Patent No. 6183968

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Lal, Preeti

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Yue, Henry

; APPLICANT: Reddy, Roopa

; APPLICANT: Guegler, Karl J.

; APPLICANT: Baughn, Mariah R.

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING

; NUMBER OF SEQUENCES: 134

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/276,531

; FILING DATE: Herewith

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/079,677

; FILING DATE: March 27, 1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Lynn E. Murry, Ph.D.


```
;
;
;   REGISTRATION NUMBER: 42,918
;   REFERENCE/DOCKET NUMBER: PA-0008 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (650) 855-0555
;   TELEFAX: (650) 845-4166
;   INFORMATION FOR SEQ ID NO: 16:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 4245 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     LIBRARY: BRAITUT21
;     CLONE: 2526432
;   US-09-276-531-16

Query Match
Best Local Similarity 43.8%; Score 39.6; DB 4; Length 4245;
Matches 171; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 357 accgtctgaagacatctctgctgctgaaggtcatcaagctaacctcttcagctgcag 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 AGCTTCACATCGGACATCGGCCCGTACACAGCGGGCGCTTCTACACCAAGCCACTCAG 189

QY 417 gtgaagcctcgcagcaatgacagagctgctagtcagatccctctgacatgctgct 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 TGGCTGGACATCCCGCGGACCTGGCGCTGTGCGCACACGTGGCTACAGAGATGCTG 249

QY 477 gatttaacacgcccctgtgtcaagacatctgtgagttccacatgaagctgaagccaa 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 CTGCCACCTGCTGGAGACGACGACCATGCGGAGGTGACAGCAGCAGGACGACACTGG 309

QY 537 gccaccatcgcattgtaacaccagtgtcaagtgtgccaccaccgctgtgtctcaagtct 586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 GGGCCCCCTCTCAACAAGACTGTCACGCCGGGACCCAGGCTTCTGCTGCTGCTTTC 369

QY 597 gccaccagcagatggaagctgcagctgcacacactgtgtcataaagctctcttcgtgaa 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 GCGCCCGTGTGCTGACCGGCGGCATCTACCCGTGCGTGTGCTGCTGCGGCGGCGG 429

QY 657 gcccttagtaagcaggtcatgaacctctctagtgcatccctctgccaatcagtgaa 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 430 GACTCGTGGAGCGCGCTCATGTGCTTCTGCGCTTACTGCGCCGAGATGTTAAGTGT 489

QY 717 cagctgtgtcccggtgatgaggtctcttc 746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 GACAAGTCCCGAGGGGAGACGTGTGATC 519

RESULT 4
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
```

```
;
;   OTHER INFORMATION: represent a, t, c or g
;   US-09-103-840A-2

Query Match
Best Local Similarity 50.0%; Score 38.4; DB 4; Length 4403765;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 465 gacatggtgctgattcaacacagccctgtgttaaacatctgtgtgagttccacatgaag 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291448 GAGGTGTGCTGCGACAGGCGCCACAGTGTGCGCATCATGTGATGATCGCGCGGAG 291389

QY 525 actgagcccaagccacatctcgcagatggaacacagtgcaagtggtgccaccgcctgtc 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291388 AACCTGGCCGAAMCCCGCAGCAAGTGGGGGCGACCGCGTGTGTGCTGACGTCACCGCC 291329

QY 585 ctcaagtactgtgtccacagccatgtggaagctgtgcacatccaaactgtctgaagctctcc 644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291328 GACGATGCGGTGCAAGATCAGCAGCACTGCGGACCAACGAGGTGCAAGCGCGGAC 291269

QY 645 ttctgtgtgaac 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291268 ATCTGTGTCAAC 291257

RESULT 5
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damej, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Epithomes
; FILE REFERENCE: 0867/ID903051
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
; US-09-130-114-2

Query Match
Best Local Similarity 43.9%; Score 38.2; DB 2; Length 1931;
Matches 163; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 331 tccctgtgtgagcagctgtgtgaacacgctctgaagacacatcatctgtgtgaagttca 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 624 tcccgctcccccgtctcgtctcccgctctgtctctcctccctcccgctcccgctcc 683

QY 391 tcacagctaacatctccacagctgcaagtcgaagccctcgccgaatgaccaggaagctgtag 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 tcccgctctcgtctcccgctctcgtctcccgctctcgtctcccgctctcgtctcccgctcc 743

QY 451 tcaagatcccccgtgaagatgtgtgtgattcaacacgcccctgtgtcaagaagatcgtgtg 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 tcccgctcccccgtctcgtctcccgctctcgtctcccgctctcgtctcccgctcccgctcc 803

QY 511 agttccacatgaagatgagggcccaagccacatctcgaatgagacacacagtgaaagtggtc 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 804 ccgtctcgtctcctcccgctctcccgctctcccgctctcgtctcctcccgctctcccgctcc 863

QY 571 ccacccgctgtgtcctcagtgactgtgtccacagccatgtggaagctgtgcgacatccaatg 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 864 tgtctcccgctcccgctcccgctctcgtctcctcccgctcccgctcccgctctcgtctcccg 923

QY 631 tgcataagctctctcctctgtgtgaagcgtttagtaagcagtgatgaacctctctagtcgc 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

| QY | Db |
|----------------------|---|
| 691 catcctggcc 701 | 924 tctccccgctactgctctctctcccgctcctgctcctcccgctctctgctctccactcc 983 |
| | |
| 984 ggccccaagctc 994 | |

RESULT 6
ME-09-904

```

US-08-804-166-7
; Sequence 16, Application US/08B04166
; Patent No. 6193972
GENERAL INFORMATION:
APPLICANT: Campbell, Robert K.
APPLICANT: Jameson, Bradford A.
APPLICANT: Chappel, Scott C.
TITLE OF INVENTION: HYBRID PROTEINS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street N.W., Ste. 300
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 22207
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,166
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011,936
FILING DATE: 20 February 1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: CAMPBELL-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ. ID NO.: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1301 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 279..1287
US-08-804-166-7

```

| | | | | |
|--------------------------|--------|----------------|-----------|--------------|
| Query Match | 2.28; | Score 37.6; | DB 4; | Length 1301; |
| Best Local Similarity | 51.88; | Pred. No. 1.1; | | |
| Matches 85; Conservative | 0; | Mismatches 79; | Indels 0; | Gaps 0 |

QY 283 cagcagcctgcgcgcctgcctcagtcagcctgcggaagaaccagccgagagatccctcgtctg 34

Db 856 CAGCTGGTGCCTGCTCCACAGGCTGCGCCCATCATTCGACCCCTGCTGGAGAGGAGG 915

QY 343 gcacccctggtbaaacccgtcctcgaagacacatcattcgttgaaggtatcaagctaca 402

Db 916 GCTGCCCGGTGTGATACCGCTAACACCACTATCTGCGCGGTACTGCGCCACCATGA 975

QY 403 tccctcagctgcagctgtagagccctccggtccaatgacacagagctg 446

Db 976 CCGCGGTGCTGCAAGGGGCTCTGCGCGCCCTGCTCAGAGGTG 1019

RESULT 7
US-08-910-991-7
; Sequence 7, Application US/08910991
; Patent No. 6104177

APPLT
APPLT

```

1 APPLICANT: Jameson, Bradford A.
2 APPLICANT: Chappel, Scott C.
3 TITLE OF INVENTION: HYBRID PROTEINS
4 NUMBER OF SEQUENCES: 22
5 CORRESPONDENCE ADDRESS:
6 ADDRESS: BROWDY AND NEIMARK
7 STREET: 419 Seventh Street N.W., Ste. 300
8 CITY: Washington
9 STATE: D.C.
10 COUNTRY: USA
11 ZIP: 22207
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.30
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/910,991
19 FILING DATE:
20 CLASSIFICATION: 530
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: 08/804,166
23 FILING DATE: 20 February 1997
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: 60/011,996
26 FILING DATE: 20 February 1996
27 ATTORNEY/AGENT INFORMATION:
28 NAME: YUN, Allen C.
29 REGISTRATION NUMBER: 37,971
30 REFERENCE/DOCKET NUMBER: CAMPBELL=2B
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (202) 628-5197
33 TELEFAX: (202) 737-3528
34 INFORMATION FOR SEQ. ID NO: 7:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 1301 base pairs
37 TYPE: nucleic acid
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 MOLECULE TYPE: cDNA
41 FEATURE:
42 NAME/KEY: CDS
43 LOCATION: 279..1287
44 OS-08-910-991-7

```

| | | | | |
|-----------------------|--------|----------------|--------|----------------|
| Query Match | 2.28; | Score 37.6; | DB 4; | Length 1301; |
| Best Local Similarity | 51.8%; | Pred. No. 1.1; | | |
| Matches | 85; | Conservative | 0; | Mismatches 79; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

[illegible]

RESULT 8
US-09-434-288-11
; Sequence 11, Application US/09434288


```

1 STREET: 7 Skyline Drive
2 City: Hawthorne
3 STATE: NY
4 COUNTRY: USA
5 ZIP: 10532
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: PatentIn Release #1.0, Version #1.30B
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/471,033
14 FILING DATE:
15 CLASSIFICATION: 530
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 08/314,594
18 FILING DATE: 09-SEP-1994
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US 08/218,018
21 FILING DATE: 23-MAR-1994
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 08/037,057
24 FILING DATE: 25-MAR-1993
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Pace, Gary M.
27 REGISTRATION NUMBER: P-40,403
28 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 919-541-8582
31 TELEFAX: 919-541-8689
32 INFORMATION FOR SEQ ID NO: 42:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 1241 base pairs
35 TYPE: nucleic acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 MOLECULE TYPE: other nucleic acid
39 DESCRIPTION: /desc = "Synthetic DNA"
40 HYPOTHEetical: NO
41 FEATURE:
42 NAME/KEY: CDS
43 LOCATION: 9..1238
44 OTHER INFORMATION:
45 OTHER INFORMATION:
46 OTHER INFORMATION:
47 OTHER INFORMATION:
48
49 /note= "Maize optimized DNA
50 sequence encoding VIF2A(a) with the Bacillus secretion signal
51 removed and the eukaryotic secretion signal inserted as
52 contained in pcIB5528"
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

| | | | | |
|-----------------------|--------------|--|-----------------|--------------|
| Query Match | 2.2% | Score 37.2; | DB 1; | Length 1241; |
| Best Local Similarity | 48.6%; | Freq. No.1.3; | | |
| Matches 102; | Conservative | 0; | Mismatches 108; | Indels 0; |
| | | | | Gaps 0; |
| QY | 353 | gaacacgctccctgaagacacatcatctgctgtaagttacacagcttaacatcttcagct | 412 | |
| | | | | |
| Db | 155 | GAAGACACTCAAGACCACACTACAAAGAAATCACCTTCAGCATAGCCGGAGCTTGAGGA | 214 | |
| | | | | |
| QY | 413 | gcaggtgaagccctctcgcaatgacacagagctgtcgtactgaagatcccccctgagactgt | 472 | |
| | | | | |
| Db | 215 | CGAGATCAAGGACCTGGAGGAGATGGAAAGATGTTCCAGCAAGACCACCTTAGCAACAG | 274 | |
| | | | | |
| QY | 473 | ggtcgtgattcaacacgccccttgytctaagacatcgttgagttccacatgacgactgagtc | 532 | |
| | | | | |
| Db | 275 | CATCATCACTCAACAAAGAAAGTGGAGCCACCAACCATCGGCTTCACAAAGAGGCTTACCGA | 334 | |
| | | | | |
| QY | 533 | ccaagccaccctccgcattgtagacaccagtgc | 562 | |
| | | | | |
| Db | 335 | GGGCAACACCATCAACAGCAGGCCATGCG | 364 | |

RESULT 11
US-08-471-044-39
: Sequence 39, Application US/08471044

```

1 Patent No. 5840868
2 GENERAL INFORMATION:
3 APPLICANT: Warren, Gregory W
4 APPLICANT: Koziel, Michael G
5 APPLICANT: Mullins, Martha A
6 APPLICANT: Nye, Gordon J
7 APPLICANT: Carr, Brian
8 APPLICANT: Desai, Nalini M
9 APPLICANT: Kostelchka, N. Kristy
10 APPLICANT: Duck, Nicholas B
11 APPLICANT: Estruch, Juan J
12 TITLE OF INVENTION: No. 5840868e1 Pesticidal Proteins and Strains
13 NUMBER OF SEQUENCES: 50
14 CORRESPONDENCE ADDRESS:
15 ADDRESSEE: CIBA-GEIGY Corporation
16 STREET: 7 Skyline Drive
17 CITY: Hawthorne
18 STATE: NY
19 COUNTRY: USA
20 ZIP: 10532
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23 COMPUTER: IBM PC compatible
24 OPERATING SYSTEM: PC-DOS/MS-DOS
25 SOFTWARE: Patent In Release #1.0, Version #1.30B
26 CURRENT APPLICATION DATA:
27 APPLICATION NUMBER: US/08/471,044
28 FILING DATE: 06-JUN-1995
29 CLASSIFICATION: 800
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/463,483
32 FILING DATE: 05-JUN-1995
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 08/314,594
35 FILING DATE: 09-SEP-1994
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 08/218,018
38 FILING DATE: 23-MAR-1994
39 PRIOR APPLICATION DATA:
40 APPLICATION NUMBER: US 08/037,057
41 FILING DATE: 25-MAR-1993
42 ATTORNEY/AGENT INFORMATION:
43 NAME: Pace, Gary M.
44 REGISTRATION NUMBER: 40,403
45 REFERENCE/DOCKET NUMBER: CGC 1695/CI/P3/DIV6 - SOLV3
46 TELECOMMUNICATION INFORMATION:
47 TELEPHONE: 919-541-8582
48 TELEFAX: 919-541-8689
49 INFORMATION FOR SEQ ID NO: 39:
50 SEQUENCE CHARACTERISTICS:
51 LENGTH: 1241 base pairs
52 TYPE: nucleic acid
53 STRANDEDNESS: single
54 TOPOLOGY: linear
55 MOLECULE TYPE: other nucleic acid
56 DESCRIPTION: /desc = "Synthetic DNA"
57 HYPOTHETICAL: NO
58 FEATURE:
59 NAME/KEY: CDS
60 LOCATION: 9..1238
61 OTHER INFORMATION:
62 OTHER INFORMATION:
63 OTHER INFORMATION: removed as contained in pcIB557"
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
90
```

| | | | | |
|---------------------------|-------|----------------|----------|-------------|
| Query Match | 2.2% | Score 37.2 | DB 2 | Length 1241 |
| Best Local Similarity | 48.6% | Pred No. 1.3 | | |
| Matches 102, Conservative | 0 | Mismatches 108 | Indels 0 | Gaps 0 |

| | | | |
|----|-----|---|-----|
| QY | 353 | gaacaccgtctcgaagacatcatcattgctcgtaaagtcatacaagctaacaatccacagt | 412 |
| | | | |
| Db | 155 | gaacgcacatcgaaccacatcgaacgaatcaccctcagatcattacccggcgcttccagga | 214 |
| | | | |

```

RESULT 12
US-08-471-044-42
? Sequence 42, Application US/08471044
? Patent No. 5840868
?
? GENERAL INFORMATION:
?
? APPLICANT: Warren, Gregory W
?
? APPLICANT: Kozielec, Michael G
?
? APPLICANT: Mullins, Martha A
?
? APPLICANT: Nye, Gordon J
?
? APPLICANT: Carr, Brian
?
? APPLICANT: Desai, Nalini M
?
? APPLICANT: Kostichka, N. Kristy
?
? APPLICANT: Duck, Nicholas B
?
? APPLICANT: Estruch, Juan J
?
? TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
?
? NUMBER OF SEQUENCES:.. 50

```

ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"

| | | | | |
|---------------------------|-------|-----------------|-----------|--------------|
| Query Match | 2.2% | Score 37.2; | DB 2; | Length 1241; |
| Best Local Similarity | 48.6% | Pred. No. 1.3; | | |
| Matches 102; Conservative | 0; | Mismatches 108; | Indels 0; | Gaps 0; |

| QY | 353 | gaacacgcctccgtagagacatcatctgcgtgaagtcatacaagctaaactccctccagct | 412 |
|----|-----|---|-----|
| Db | 155 | GAAGGACACTCAAGACCCANCTACAAAGGAGATACCTTACGCTACGCCGACCTTCGAGGA | 214 |
| QY | 413 | gcaggtaagagccctccgccaatgaccagagcgtctgataaagaatcccccctgagacatggt | 472 |
| Db | 215 | CGAGATCAAGGACCTGAAAGGAGATCGAACAAAGATGTTGCAAGACCAACTGACCAACAG | 274 |
| QY | 473 | gagctgagatcaacagcccccctggtcctaagaccatctgtgagttccacatgtagagactgagc | 532 |
| Db | 275 | CATCATCACTTACCAAGAACGTGGAGCCACCACCACCATCGGCTTCAACAAAGAGCCTGACCGA | 334 |
| QY | 533 | ccaagcaccacatccgcatgagcagacatgagc | 562 |
| Db | 335 | GGGCAACACCATCAACAGCGAGCGCCATGGC | 364 |

RESULT 13
 US-08-463-483A-39
 Sequence 39, Application US/08463483A
 Patent No. 5849870
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Kozien, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalini M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/463,483A
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993

```

ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal
OTHER INFORMATION: removed as contained in pcib5527"
US-08-463-483A-39

```

```

Query Match
Best Local Similarity 48.6%; Score 37.2; DB 2; Length 1241;
Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

```

```

QY 353 gaacacgctcgaagcaccatcatcgtgctgaagtcatacaccatccacagct 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 155 GAAGCATCTCAAGACCACTCAAGAGATCACTTCAAGCGGCGACCTTGAGGA 214
QY 413 gcaggtgaagccctcgcccaatgaccagagctgtctgaagaatcccccctgacatggt 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 215 CGAGATCAAGACCTGAAGAGATCGACAGATGTTGACAAAGACCAACCTGAGCAACAG 274
QY 473 ggcgtgattcaacagcccccctggtcaagaccatcgtgtgagttccacatgacagcagtc 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275 CATCATCACTCAAGAAAGAGTGAGCCCAACCATCGGCTTCAACAAGAGCCCTGACCGA 334
QY 533 ccaagccaccatccgcatgacacacagtcg 562
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 335 GGGCAACACCATCAACAGCGACGCCCATGGC 364

```

```

RESULT 14
US-08-463-483A-42
Sequence 42, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion sig
OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
OTHER INFORMATION: contained in pcib5528"
US-08-463-483A-42

```

```

Query Match
Best Local Similarity 48.6%; Score 37.2; DB 2; Length 1241;
Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

```

```

QY 353 gaacacgctcgaagcaccatcatcgtgctgaagtcatacaccatccacagct 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 155 GAAGCATCTCAAGACCACTCAAGAGATCACTTCAAGCGGCGACCTTGAGGA 214
QY 413 gcaggtgaagccctcgcccaatgaccagagctgtctgaagaatcccccctgacatggt 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 215 CGAGATCAAGACCTGAAGAGATCGACAGATGTTGACAAAGACCAACCTGAGCAACAG 274
QY 473 ggcgtgattcaacagcccccctggtcaagaccatcgtgtgagttccacatgacagcagtc 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275 CATCATCACTCAAGAAAGAGTGAGCCCAACCATCGGCTTCAACAAGAGCCCTGACCGA 334
QY 533 ccaagccaccatccgcatgacacacagtcg 562
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 335 GGGCAACACCATCAACAGCGACGCCCATGGC 364

```

```

RESULT 15
US-08-471-046A-39
Sequence 39, Application US/08471046A
Patent No. 5866326
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy

```

APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: Method for Isolating Vegetative Insecticidal
 TITLE OF INVENTION: Protein Genes
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 586632artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.308
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,046A
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1241 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "synthetic DNA"
 HYPOTHETICAL: NO
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 9..1238
 OTHER INFORMATION: /note="Maize optimized DNA"
 OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion signal
 OTHER INFORMATION: removed as contained in pcib5527"
 US-08-471-046A-39

Query Match 2.2%; Score 37.2; DB 2; Length 1241;
 Best Local Similarity 48.6%; Pred. No. 1.3;
 Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 353 gaacacgctcgaagacacatcatgtgtgaaggtcatcacagctaacatcctcagct 412
 DB 155 GAACGACATCAAGACCACTACAGAGGATCACTTACGATAGCCGCGACGCTTCGAGGA 214
 QY 413 gcaagtgaaagccctggyccaatgacagagagctgctagtcgaagcccttgagacatggt 472
 DB 215 CGAGATCAAGAGACCTTAAGAGATGCAACAAGATGTTGACACAGACCACTTGAGCAGAG 274
 QY 473 ggcctgattcaacacgcccctgtgcaagacatcgtgaggtccacatgacgactgagac 532
 DB 275 CATCATCACCTTCAAGAGAGCGTGGAGCCGACACCATCGGCTTCAACAGAGCGCTGACGGA 334

QY 533 ccaagccaccatccgcatgtgacacccagtyc 562
 DB 335 GGGCAACACCATCAACAGCGAGCGCATGSC 364

Search completed: September 16, 2002, 03:26:41
 Job time: 9324 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:31:11 : Search time 390.66 Seconds
(without alignments)
1384.396 Million cell updates/sec

Title: US-09-700-770-4

Perfect score: 315

Sequence: 1 taacactgcactcagattt.....tgtgtattactactatgc 315

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N.Geneseq_032802:*

```
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 315 | 100.0 | 315 | 21 | AAZ29725 |
| 2 | 21 | 6.7 | 133894 | 17 | AAT13635 |
| 3 | 18 | 5.7 | 9620 | 18 | AAT45424 |
| 4 | 18 | 5.7 | 23934 | 22 | ABA19145 |
| 5 | 18 | 5.7 | 23934 | 22 | AAL36171 |
| 6 | 18 | 5.7 | 23934 | 22 | AAL36179 |
| 7 | 18 | 5.7 | 23934 | 22 | AAL04522 |
| 8 | 18 | 5.7 | 23934 | 22 | AAS28343 |
| 9 | 18 | 5.7 | 23934 | 22 | AAK71442 |

| | | | | | | |
|----|----|-----|-------|----|----------|----------------------|
| 10 | 17 | 5.4 | 699 | 22 | AAH92576 | Human inflammatory |
| 11 | 17 | 5.4 | 726 | 22 | AAH04423 | Human cDNA clone (|
| 12 | 17 | 5.4 | 842 | 21 | AAZ36779 | Nucleotide sequenc |
| 13 | 17 | 5.4 | 1239 | 22 | AAH94373 | Human foetal cDNA, |
| 14 | 17 | 5.4 | 1959 | 22 | ABA45382 | Human breast cell |
| 15 | 17 | 5.4 | 1959 | 22 | ABA55872 | Human foetal liver |
| 16 | 17 | 5.4 | 1959 | 22 | ABA25546 | Probe #4012 for ge |
| 17 | 17 | 5.4 | 1959 | 22 | AAK04090 | Human brain expres |
| 18 | 17 | 5.4 | 1959 | 22 | AAK29571 | Human bone marrow |
| 19 | 17 | 5.4 | 1959 | 22 | AAI14146 | Probe #4079 for ge |
| 20 | 17 | 5.4 | 1959 | 22 | AAI35531 | Probe #4217 used t |
| 21 | 17 | 5.4 | 1959 | 22 | AAI03994 | t Arabidopsis thalia |
| 22 | 17 | 5.4 | 2385 | 21 | AAC51459 | Arabidopsis thalia |
| 23 | 17 | 5.4 | 2388 | 21 | AAC36512 | Arabidopsis thalia |
| 24 | 17 | 5.4 | 2879 | 23 | ABL27076 | Drosophila melanog |
| 25 | 17 | 5.4 | 3008 | 23 | ABL21896 | Drosophila melanog |
| 26 | 17 | 5.4 | 3201 | 23 | ABL02390 | Drosophila melanog |
| 27 | 17 | 5.4 | 3241 | 23 | ABL20736 | Drosophila melanog |
| 28 | 17 | 5.4 | 3424 | 23 | ABL21898 | Drosophila melanog |
| 29 | 17 | 5.4 | 4060 | 23 | ABL07106 | Drosophila melanog |
| 30 | 17 | 5.4 | 4146 | 17 | AAT09026 | Arabidopsis thalia |
| 31 | 17 | 5.4 | 4645 | 22 | AAS35982 | Human cardiovascular |
| 32 | 17 | 5.4 | 4646 | 22 | AAS35981 | Human cardiovascular |
| 33 | 17 | 5.4 | 5406 | 22 | AAF83912 | cDNA sequence of p |
| 34 | 17 | 5.4 | 5407 | 23 | ABL18530 | Drosophila melanog |
| 35 | 17 | 5.4 | 5408 | 21 | AAAI0226 | Human PCNA-1 splic |
| 36 | 17 | 5.4 | 5532 | 22 | AAF83913 | cDNA sequence of p |
| 37 | 17 | 5.4 | 5534 | 21 | AAAI0227 | Human PCNA-1 splic |
| 38 | 17 | 5.4 | 6236 | 23 | ABL23206 | Drosophila melanog |
| 39 | 17 | 5.4 | 8031 | 23 | ABL20966 | Drosophila melanog |
| 40 | 17 | 5.4 | 9931 | 23 | ABL07088 | Drosophila melanog |
| 41 | 17 | 5.4 | 12151 | 21 | AAZ36757 | DNA sequence encod |
| 42 | 17 | 5.4 | 15866 | 22 | AAK90892 | Human digestive sy |
| 43 | 17 | 5.4 | 15862 | 22 | AAL04554 | Human reproductive |
| 44 | 17 | 5.4 | 19802 | 22 | AAK90894 | Human digestive sy |
| 45 | 17 | 5.4 | 21403 | 22 | AAI04555 | Human reproductive |

ALIGNMENTS

| | |
|---|------------------------|
| RESULT 1 | |
| AAZ29725 | standard; DNA; 315 BP. |
| AAZ29725; | |
| 27-MAR-2000 | (first entry) |
| Human Lung specific gene-2. | |
| Human Lung specific gene-2. | |
| Lung Specific Gene; LSG; human; diagnostic marker; | |
| prognosticate; lung cancer; diagnosis; ds. | |
| Homo sapiens. | |
| WO9960160-A1. | |
| 25-NOV-1999. | |
| 12-MAY-1999; | 99WO-US10344. |
| 21-MAY-1998; | 98US-0086212. |
| (DIAD-) DIADEXUS LLC. | |
| Yang F, Macina RA, Sun Y; | |
| WPI, 2000-116320/10. | |
| A new method for diagnosing, monitoring and staging lung cancer | |
| Example 1; Page 35; 40pp; English. | |

XX The present sequence is a lung specific gene (LSG) from human
 CC clone ID 1472038. The LSG has high level of tissue specificity for lungs
 CC and is overexpressed in cancerous tissues. The sequence serves as a
 CC diagnostic marker for detecting, monitoring, staging and prognosticating
 CC lung cancer. The diagnosis involves comparing levels of LSG in samples
 CC obtained from patient and normal control.
 XX
 SQ Sequence 315 BP; 108 A; 58 C; 49 G; 100 T; 0 other;

Query Match
 Best Local Similarity 100.0%; Score 315; DB 21; Length 315;
 Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taaacactgactcagattttaagaataactttgagaagaataagaataatgattcgtt 60
 Db 1 taaacactgactcagattttaagaataactttgagaagaataagaataatgattcgtt 60
 QY 61 tctccacacttaagtatactcttagagatctacagctccctcttagaggagacatacaa 120
 Db 61 tctccacacttaagtatactcttagagatctacagctccctcttagaggagacatacaa 120
 QY 121 gtccagttgtgtccttctgttgagtcaccacttatattcaagtagtgactacaatt 180
 Db 121 gtccagttgtgtccttctgttgagtcaccacttatattcaagtagtgactacaatt 180
 QY 181 ttgaataatagttgcacacataaactgtgagttatgagaacatcagtagaagaataa 240
 Db 181 ttgaataatagttgcacacataaactgtgagttatgagaacatcagtagaagaataa 240
 QY 241 caacattccaccccttaacagagatcattacttgcacactcagtagaagaataattgtcattgt 300
 Db 241 caacattccaccccttaacagagatcattacttgcacactcagtagaagaataattgtcattgt 300
 QY 301 attactactatgc 315
 Db 301 attactactatgc 315

RESULT 2
 AAT13635/c
 ID AAT13635 standard; DNA: 133894 BP.
 AC AAT13635;
 AT 03-SEP-1996 (first entry)
 DE ACNPV genomic DNA clone 6.
 KW Autographa californica nuclear polyhedrosis virus clone 6;
 KW disruption; non-essential gene; heterologous protein production;
 KW expression vector; baculovirus; ss.
 OS Autographa californica nuclear polyhedrosis virus clone 6.
 PN WO9601320-A2.
 PD 18-JAN-1996.
 PF 30-JUN-1995; 95WO-1B00578.
 PR 04-JUL-1994; 94GB-0013420.
 PA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
 PI Ayres M, Bishop D, Possee R;
 DR WPI: 1996-087670/09.
 DR GENBANK: L22858.
 PT Autographa californica nuclear polyhedrosis virus complete genome
 PT sequence - useful in the prodn. of vectors for enhanced

PT heterologous protein expression, such as interleukin(s),
 PT interferon(s) and neurotoxin(s)
 XX
 PS Disclosure; Page 90-186; 122pp; English.
 XX
 CC The complete nucleotide sequence of the genome of clone 6 of the
 CC baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV)
 CC has been determined. The sequence is taken from the Genbank record
 CC L22858. The patent specification claims a polynucleotide selected from
 CC open reading frames (ORFs 13, 20, 22-26, 28-30, 32, 38, 41-46, 50-60,
 CC 62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,
 CC 140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by
 CC the patentees. See T13636-731. Expression vectors congt. the complete
 CC genomic sequence of AcNPV, with the exception that at least one non-
 CC essential ORF is disrupted or replaced are useful for the synthesis of
 CC heterologous proteins.
 XX
 SQ Sequence 133894 BP; 39195 A; 27151 C; 27347 G; 40201 T; 0 other;

Query Match
 Best Local Similarity 6.7%; Score 21; DB 17; Length 133894;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 aagaataactttgagaat 41
 Db 114255 AACAAATTAATTTGAGAAAT 114235

RESULT 3
 AAT45424
 ID AAT45424 standard; cDNA: 9620 BP.
 AC AAT45424;
 AT 30-MAR-1998 (first entry)
 DE Murine Ataxia-telangiectasia ATM gene.
 KW Ataxia-telangiectasia; A-T; mutated; ATM; 11q22-23; signal transduction;
 KW DNA damage; cell cycle control; screening; gene therapy; catalytic;
 KW Phosphatidylinositol-3 kinase; PI-3; cancer predisposition; carrier;
 KW mouse; ss.
 OS Mus musculus.
 PN WO9636695-A1.
 PD 21-NOV-1996.
 PF 16-MAY-1996; 96WO-US07040.
 PR 28-JUL-1995; 95US-0508836.
 PR 16-MAY-1995; 95US-0441822.
 PR 21-JUN-1995; 95US-0493092.
 PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Collins FS, Shiloh Y, Tagle DA;
 DR P-PSDB; AAM36178.
 PT New gene ATM implicated in ataxia-telangiectasia and related protein
 PT carriers
 PT - useful in screening methods, partic. for identifying disease
 PS Example 4; Page 90-95; 127pp; English.

XX This sequence represents the mouse ATM gene, mutations in which cause
CC ataxia-telangiectasia (A-T), a progressive genetic disorder affecting
CC the central nervous and immune systems. In the human ATM gene, there is
CC extensive alternate splicing at the 5' UTR of the ATM transcript giving
CC rise to 12 different 5'UTRs (the longest is given in AAT45421). The
CC first exon in the human sequence is designated 1b. An alternative leader
CC exon, 1a is given in AAT45422 and the 3' UTR is in AAT45420. The ATM
CC gene, located in humans at chromosome 11q22-23, is probably involved in a
CC novel signal transduction system that links DNA damage surveillance to
CC cell cycle control. This murine sequence is located in chromosome 9, band
CC 9C. The human ATM gene product (AAM07655) has a highly conserved
CC C-terminal region showing high sequence homology to the catalytic domain
CC of phosphatidylinositol-3 kinases. A-T mutations affect a variety of
CC tissues and lead to cancer predisposition. Identification of A-T
CC carriers, by analysis at nucleic acid or protein levels, allows better
CC supervision and treatment of such subjects who are at increased risk of
CC developing cancer and are particularly sensitive to radiation. The
CC transgenic animals and transformed cells are useful as models of the
CC human disease. Also viral vectors expressing the ATM protein can be used
CC in gene therapy of A-T.

XX
SQ Sequence 9620 BP; 2898 A; 1931 C; 2174 G; 2617 T; 0 other;

Query Match 5.7%; Score 18; DB 18; Length 9620;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 tctcttagagactctacag 97
|||||

Db 6071 tctcttagagactctacag 6088

RESULT 4

ABAI9145

ID ABAI9145 standard; DNA; 23934 BP.

XX
AC ABAI9145;

XX
DT 23-JAN-2002 (first entry)

XX
DE Human nervous system related polynucleotide SEQ ID NO 11476.

XX
KW Human; nocitropic; neuroprotective; cytosolic; dermatological; virocidic;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerray;
KW antiparkinsonian; antistickling; antianaemic; antiarthritis; cancer;
KW antitubercular; hepatotrophic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX
OS Homo sapiens.

XX
PN WO200159063-A2.

XX
PD 16-AUG-2001.

XX
PF 17-JAN-2001; 2001WO-US01334.

XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229511.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.

PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2000US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-541565/60.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Disclosure; SEQ ID NO 11476; 1701bp + Sequence Listing; English.
 PS
 XX The invention relates to novel genes (ABR11004-ABR21534) and proteins
 CC (ABR14678-ABR18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 23934 BP; 7785 A; 4761 C; 4365 G; 7023 T; 0 other;
 Query Match 5.7%; Score 18; DB 22; Length 23934;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 172 ctacaatttgaata 189
 DB 10877 ctacaatttgaata 10894
 ||||||||||||||||
 RESULT 5
 AAL36171/C
 ID AAL36171 standard; DNA; 23934 BP.
 XX
 AC AAL36171;
 XX
 DT 08-JAN-2002 (first entry)
 XX
 DE Human musculoskeletal system related polynucleotide SEQ ID NO 2536.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cariant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein;
 KW musculoskeletal system; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO20015367-A1.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01338.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 11-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.

PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-451937/48.
DR
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the musculoskeletal system including
PT musculoskeletal cancers and also for testing and detection e.g.
PT diagnosis -
PS
PS Example 2; SEQ ID NO 2536; 781pp + Sequence listing; English.
XX
XX The invention relates to novel genes (A134669-A137666) and proteins
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (anti)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic
CC anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WHO at ftp.who.int/pub/publistned_pcc_sequences.
XX
XX Sequence 23934 BP; 7023 A; 4365 C; 4761 G; 7785 T; 0 other;

Query Match 5.7%; Score 18; DB 22; Length 23934;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 ctacaatttggaaaata 189
|||||
Db 13058 CTACAAATTGAAAATA 13041

RESULT 6

AAL36179

ID AAL36179 standard; DNA: 23934 BP.

XX AAL36179;

XX 08-JAN-2002 (first entry)

XX Human musculoskeletal system related polynucleotide SEQ ID NO 2544.

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;

KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antilicer;

KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;

KW cardiant; gene therapy; Cancer; immune disorder; cardiovascular disorder;

KW neurological disease; infection; human; secreted protein;

KW musculoskeletal system; ds.

XX Homo sapiens.

XX WO200155367-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US0138.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226861.

XX 22-AUG-2000; 2000US-0226868.

XX 23-AUG-2000; 2000US-0227182.

XX 30-AUG-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

XX 01-SEP-2000; 2000US-0229343.

XX 01-SEP-2000; 2000US-0229344.

XX 17-NOV-2000; 2000US-0249214.

XX 17-NOV-2000; 2000US-0249215.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 13-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 20-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 01-NOV-2000; 2000US-0241826.

PR 08-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

| | | |
|----|--|---------------|
| XX | AAI04522; | |
| AC | | |
| XX | | |
| DT | 21-NOV-2001 | (first entry) |
| XX | | |
| DE | Human reproductive system related antigen DNA SEQ ID NO: 7210. | |
| XX | | |
| KW | Human; reproductive system related antigen; reproductive system disorder | |
| KW | cancer; gene therapy; ds. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| PN | WO200155320-A2. | |
| XX | | |
| PD | 02-AUG-2001. | |
| XX | | |
| PF | 17-JAN-2001; 2001WO-US01339. | |
| XX | | |
| PR | 31-JAN-2000; 2000US-0179065. | |
| PR | 04-FEB-2000; 2000US-0180628. | |
| PR | 24-FEB-2000; 2000US-0184664. | |
| PR | 02-MAR-2000; 2000US-0186350. | |
| PR | 16-MAR-2000; 2000US-0189874. | |
| PR | 17-MAR-2000; 2000US-0190076. | |
| PR | 18-APR-2000; 2000US-0198123. | |
| PR | 19-MAY-2000; 2000US-0205515. | |
| PR | 07-JUN-2000; 2000US-0209467. | |
| PR | 28-JUN-2000; 2000US-0214886. | |
| PR | 30-JUN-2000; 2000US-0215135. | |
| PR | 07-JUL-2000; 2000US-0216647. | |
| PR | 07-JUL-2000; 2000US-0216880. | |
| PR | 11-JUL-2000; 2000US-0217487. | |
| PR | 11-JUL-2000; 2000US-0217496. | |
| PR | 14-JUL-2000; 2000US-0218290. | |
| PR | 26-JUL-2000; 2000US-0220963. | |
| PR | 26-JUL-2000; 2000US-0220964. | |
| PR | 14-AUG-2000; 2000US-0224518. | |
| PR | 14-AUG-2000; 2000US-0224519. | |
| PR | 14-AUG-2000; 2000US-0225213. | |
| PR | 14-AUG-2000; 2000US-0225214. | |
| PR | 14-AUG-2000; 2000US-0225266. | |
| PR | 14-AUG-2000; 2000US-0225267. | |
| PR | 14-AUG-2000; 2000US-0225268. | |
| PR | 14-AUG-2000; 2000US-0225270. | |
| PR | 14-AUG-2000; 2000US-0225447. | |
| PR | 14-AUG-2000; 2000US-0225477. | |
| PR | 14-AUG-2000; 2000US-0225757. | |
| PR | 14-AUG-2000; 2000US-0225758. | |
| PR | 14-AUG-2000; 2000US-0225759. | |
| PR | 18-AUG-2000; 2000US-0226279. | |
| PR | 22-AUG-2000; 2000US-0226681. | |
| PR | 22-AUG-2000; 2000US-0226686. | |
| PR | 22-AUG-2000; 2000US-0227182. | |
| PR | 23-AUG-2000; 2000US-0227009. | |
| PR | 30-AUG-2000; 2000US-0228924. | |
| PR | 01-SEP-2000; 2000US-0229287. | |
| PR | 01-SEP-2000; 2000US-0229343. | |
| PR | 01-SEP-2000; 2000US-0229344. | |
| PR | 01-SEP-2000; 2000US-0229345. | |
| PR | 05-SEP-2000; 2000US-0229509. | |
| PR | 05-SEP-2000; 2000US-0229513. | |
| PR | 06-SEP-2000; 2000US-0230437. | |
| PR | 06-SEP-2000; 2000US-0230438. | |
| PR | 08-SEP-2000; 2000US-0231242. | |
| PR | 08-SEP-2000; 2000US-0231243. | |
| PR | 08-SEP-2000; 2000US-0231244. | |
| PR | 08-SEP-2000; 2000US-0231413. | |
| PR | 08-SEP-2000; 2000US-0231414. | |
| PR | 08-SEP-2000; 2000US-0232080. | |
| PR | 08-SEP-2000; 2000US-0232081. | |
| PR | 12-SEP-2000; 2000US-0231968. | |
| PR | 14-SEP-2000; 2000US-0232397. | |
| PR | 14-SEP-2000; 2000US-0232398. | |
| PR | 14-SEP-2000; 2000US-0232399. | |

```
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0246417.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.

PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM.
XX DR WPI, 2001-465570/50.
XX
XX PT Isolated nucleic acid molecule encoding a reproductive system antigen
XX PT is used in preventing, treating or ameliorating a medical condition -
XX PS Disclosure; SEQ ID NO 7210; 1297pp + Sequence Listing; English.
XX
XX CC The present invention provides the protein and coding sequences of a
XX CC number of human reproductive system related antigens. These can be used
XX CC in the prevention and treatment of reproductive system disorders,
XX CC including cancer. The present sequence is a genomic sequence encoding a
XX CC protein of the invention.
SO Sequence 23934 BP; 7023 A; 4365 C; 4761 G; 7785 T; 0 other;

Query Match 5.7%; Score 18; DB 22; Length 23934;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ctcaaatTTGaaata 189
DB 13058 CTACAAATTTGAAATA 13041

RESULT 8
AAS28343
ID AAS28343 standard; DNA; 23934 BP.
XX
XX AAS28343;
AC
XX
XX
DT 07-NOV-2001 (first entry)
XX
XX
DE Genomic sequence #183 encoding for novel human respiratory antigen.
XX
XX Human; respiratory antigen; respiratory disorder; throat disorder;
XX KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
XX KW anti allergic; anti asthmatic; anti inflammatory; olfactory;
XX KW respiratory active; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO20015448-A1.
XX
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01333.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205415.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
```


| | | |
|----|-------------|----------------|
| PR | 11-JUL-2000 | 200005-0217487 |
| PR | 11-JUL-2000 | 200005-0217496 |
| PR | 14-JUL-2000 | 200005-0218290 |
| PR | 26-JUL-2000 | 200005-0218963 |
| PR | 26-JUL-2000 | 200005-0220964 |
| PR | 14-AUG-2000 | 200005-0224519 |
| PR | 14-AUG-2000 | 200005-0224518 |
| PR | 14-AUG-2000 | 200005-0225513 |
| PR | 14-AUG-2000 | 200005-0225514 |
| PR | 14-AUG-2000 | 200005-0225567 |
| PR | 14-AUG-2000 | 200005-0225566 |
| PR | 14-AUG-2000 | 200005-0225758 |
| PR | 18-AUG-2000 | 200005-0226679 |
| PR | 22-AUG-2000 | 200005-0226681 |
| PR | 22-AUG-2000 | 200005-0226688 |
| PR | 23-AUG-2000 | 200005-0227082 |
| PR | 30-AUG-2000 | 200005-0227089 |
| PR | 01-SEP-2000 | 200005-0229287 |
| PR | 01-SEP-2000 | 200005-0229344 |
| PR | 01-SEP-2000 | 200005-0229343 |
| PR | 01-SEP-2000 | 200005-0229509 |
| PR | 05-SEP-2000 | 200005-0229513 |
| PR | 06-SEP-2000 | 200005-0230437 |
| PR | 06-SEP-2000 | 200005-0230438 |
| PR | 08-SEP-2000 | 200005-0232081 |
| PR | 08-SEP-2000 | 200005-0231242 |
| PR | 08-SEP-2000 | 200005-0231243 |
| PR | 08-SEP-2000 | 200005-0231244 |
| PR | 08-SEP-2000 | 200005-0231413 |
| PR | 08-SEP-2000 | 200005-0231414 |
| PR | 14-SEP-2000 | 200005-0232398 |
| PR | 14-SEP-2000 | 200005-0232399 |
| PR | 14-SEP-2000 | 200005-0232400 |
| PR | 14-SEP-2000 | 200005-0232401 |
| PR | 14-SEP-2000 | 200005-0233063 |
| PR | 14-SEP-2000 | 200005-0233064 |
| PR | 21-SEP-2000 | 200005-0234223 |
| PR | 21-SEP-2000 | 200005-0234225 |
| PR | 25-SEP-2000 | 200005-0234274 |
| PR | 25-SEP-2000 | 200005-0234498 |
| PR | 26-SEP-2000 | 200005-0234584 |
| PR | 27-SEP-2000 | 200005-0235634 |
| PR | 27-SEP-2000 | 200005-0235636 |
| PR | 29-SEP-2000 | 200005-0236127 |
| PR | 29-SEP-2000 | 200005-0236167 |
| PR | 29-SEP-2000 | 200005-0236368 |
| PR | 29-SEP-2000 | 200005-0236369 |
| PR | 02-OCT-2000 | 200005-0236670 |
| PR | 02-OCT-2000 | 200005-0237037 |
| PR | 02-OCT-2000 | 200005-0237038 |
| PR | 02-OCT-2000 | 200005-0237039 |
| PR | 13-OCT-2000 | 200005-0239335 |
| PR | 13-OCT-2000 | 200005-0239337 |
| PR | 20-OCT-2000 | 200005-0240960 |
| PR | 20-OCT-2000 | 200005-0241212 |
| PR | 20-OCT-2000 | 200005-0241785 |
| PR | 20-OCT-2000 | 200005-0241786 |
| PR | 20-OCT-2000 | 200005-0241807 |
| PR | 20-OCT-2000 | 200005-0241808 |

| | | |
|----|--|-----------------|
| CC | 20-OCT-2000; | 2000US-0241809. |
| PR | 20-OCT-2000; | 2000US-0241826. |
| PR | 01-NOV-2000; | 2000US-0244617. |
| PR | 08-NOV-2000; | 2000US-0246474. |
| PR | 08-NOV-2000; | 2000US-0246475. |
| PR | 08-NOV-2000; | 2000US-0246476. |
| PR | 08-NOV-2000; | 2000US-0246477. |
| PR | 08-NOV-2000; | 2000US-0246478. |
| PR | 08-NOV-2000; | 2000US-0246523. |
| PR | 08-NOV-2000; | 2000US-0246524. |
| PR | 08-NOV-2000; | 2000US-0246525. |
| PR | 08-NOV-2000; | 2000US-0246526. |
| PR | 08-NOV-2000; | 2000US-0246527. |
| PR | 08-NOV-2000; | 2000US-0246532. |
| PR | 08-NOV-2000; | 2000US-0246538. |
| PR | 08-NOV-2000; | 2000US-0246609. |
| PR | 08-NOV-2000; | 2000US-0246610. |
| PR | 08-NOV-2000; | 2000US-0246611. |
| PR | 08-NOV-2000; | 2000US-0246613. |
| PR | 17-NOV-2000; | 2000US-0249207. |
| PR | 17-NOV-2000; | 2000US-0249208. |
| PR | 17-NOV-2000; | 2000US-0249209. |
| PR | 17-NOV-2000; | 2000US-0249210. |
| PR | 17-NOV-2000; | 2000US-0249211. |
| PR | 17-NOV-2000; | 2000US-0249212. |
| PR | 17-NOV-2000; | 2000US-0249213. |
| PR | 17-NOV-2000; | 2000US-0249214. |
| PR | 17-NOV-2000; | 2000US-0249215. |
| PR | 17-NOV-2000; | 2000US-0249216. |
| PR | 17-NOV-2000; | 2000US-0249217. |
| PR | 17-NOV-2000; | 2000US-0249218. |
| PR | 17-NOV-2000; | 2000US-0249244. |
| PR | 17-NOV-2000; | 2000US-0249245. |
| PR | 17-NOV-2000; | 2000US-0249264. |
| PR | 17-NOV-2000; | 2000US-0249265. |
| PR | 17-NOV-2000; | 2000US-0249297. |
| PR | 17-NOV-2000; | 2000US-0249299. |
| PR | 17-NOV-2000; | 2000US-0249300. |
| PR | 01-DEC-2000; | 2000US-0250160. |
| PR | 01-DEC-2000; | 2000US-0250391. |
| PR | 05-DEC-2000; | 2000US-0251030. |
| PR | 05-DEC-2000; | 2000US-0251988. |
| PR | 05-DEC-2000; | 2000US-0256719. |
| PR | 06-DEC-2000; | 2000US-0251479. |
| PR | 08-DEC-2000; | 2000US-0251856. |
| PR | 08-DEC-2000; | 2000US-0251868. |
| PR | 08-DEC-2000; | 2000US-0251869. |
| PR | 08-DEC-2000; | 2000US-0251989. |
| PR | 11-DEC-2000; | 2000US-0254097. |
| PR | 05-JAN-2001; | 2001US-0259678. |
| XX | (HUMA-) HUMAN GENOME SCI INC. | |
| PA | | |
| PI | Rosen CA, Barash SC, Ruben SM; | |
| XX | WPI; 2001-476224/51. | |
| DR | | |
| XX | | |
| PT | Isolated polypeptide for treating, preventing and/or prognosing | |
| PT | disorders related to the respiratory system including respiratory | |
| XX | cancers and also for testing and detection e.g. diagnosis - | |
| XX | | |
| XX | Disclosure; SED ID No 777; 546pp: English. | |
| PS | | |
| CC | The present invention relates to the isolation of novel human | |
| CC | respiratory antigens (AAU17685-AAU17975), and cDNA and genomic | |
| CC | sequences encoding for these polypeptides. The sequences of the | |
| CC | invention are useful for preventing, treating and/or prognosing | |
| CC | disorders related to the respiratory system including throat | |
| CC | disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis), | |
| CC | lung disorders (e.g. pneumonia, allergic disorders e.g. asthma, | |
| CC | pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of | |
| CC | the respiratory tissues e.g. lung cancer. The polynucleotide sequences | |

CC of the invention are useful in gene therapy and antisense therapy.
CC AAS28161-AAS28764 represent genomic sequences encoding for novel
CC human respiratory antigens.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 23934 BP; 7785 A; 4761 C; 4365 G; 7023 T; 0 other;

Query Match 5.7%; Score 18; DB 22; Length 23934;
Best Local Similarity 100.0%; Pred No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 172 ctacaatttgaataa 189
|||||
Db 10877 ctacaatttgaataa 10894

RESULT 9
AAK71442/c
ID AAK71442 standard; DNA; 23934 BP.
XX
AC AAK71442;
XX

DT 06-NOV-2001 (first entry)
XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26254.
XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX

OS Homo sapiens.
XX

PN WO200157182-A2.
XX

PD 09-AUG-2001.
XX

PF 17-JAN-2001; 2001MO-US01354.
XX

PR 31-JAN-2000; 2000US-0179065.
XX

PR 04-FEB-2000; 2000US-0180628.
XX

PR 24-FEB-2000; 2000US-0184664.
XX

PR 02-MAR-2000; 2000US-0186350.
XX

PR 16-MAR-2000; 2000US-0189874.
XX

PR 17-MAR-2000; 2000US-0190076.
XX

PR 18-APR-2000; 2000US-0198123.
XX

PR 19-MAY-2000; 2000US-0205515.
XX

PR 07-JUN-2000; 2000US-0209467.
XX

PR 28-JUN-2000; 2000US-0214886.
XX

PR 30-JUN-2000; 2000US-0215135.
XX

PR 07-JUL-2000; 2000US-0216647.
XX

PR 07-JUL-2000; 2000US-0216880.
XX

PR 11-JUL-2000; 2000US-0217487.
XX

PR 11-JUL-2000; 2000US-0217496.
XX

PR 14-JUL-2000; 2000US-0218290.
XX

PR 26-JUL-2000; 2000US-0220963.
XX

PR 26-JUL-2000; 2000US-0220964.
XX

PR 14-AUG-2000; 2000US-0224518.
XX

PR 14-AUG-2000; 2000US-0224519.
XX

PR 14-AUG-2000; 2000US-0225213.
XX

PR 14-AUG-2000; 2000US-0225214.
XX

PR 14-AUG-2000; 2000US-0225266.
XX

PR 14-AUG-2000; 2000US-0225267.
XX

PR 14-AUG-2000; 2000US-0225268.
XX

PR 14-AUG-2000; 2000US-0225270.
XX

PR 14-AUG-2000; 2000US-0225447.
XX

PR 14-AUG-2000; 2000US-0225757.
XX

PR 14-AUG-2000; 2000US-0225758.
XX

PR 14-AUG-2000; 2000US-0225759.
XX

PR 18-AUG-2000; 2000US-0226279.
XX

PR 22-AUG-2000; 2000US-0226681.
XX

PR 22-AUG-2000; 2000US-0226686.
XX

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234423.
PR 21-SEP-2000; 2000US-0234424.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure: SEQ ID NO 26254; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 23934 BP; 7023 A; 4365 C; 4761 G; 7785 T; 0 other;

Query Match 5.7%; Score 18; DB 22; Length 23934;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 ctcaaatatttgaataa 189
DB 13058 CTACAAATTTTGAAATA 13041

RESULT 10
AAH92576
ID AAH92576 standard; DNA; 699 BP.
XX
XX AAH92576;
XX
XX
XX 09-OCT-2001 (first entry)
XX
XX Human inflammatory bowel disease related gene fragment IGR1276a.
DE
XX
XX Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;
KW single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;
KW chromosome 5q31-33; forensic test; gene therapy; ds.
XX
XX Homo sapiens.
OS
XX WO200142511-A2.
XX
XX 14-JUN-2001.
XX
XX 11-DEC-2000; 2000WO-US33632.
XX
XX 10-DEC-1999; 99US-0170257.
XX
XX 10-APR-2000; 2000US-0196046.
XX
XX (MHED) WHITEHEAD INST BIOMEDICAL RES.
PA (ELI-) ELIPISTIS BIOTHERAPEUTICS CORP.
XX
XX Daly M, Hudson TJ, Lander ES, Rieux J, Simionovitch K;
XX
XX WPI; 2001-367874/38.
XX
XX
XX Testing for the presence of polymorphisms associated with inflammatory
PT bowel disease, using a hybridization assay -
XX
XX
XX Disclosure: Page 257; 463bp; English.
XX
XX
XX The present invention describes a method for detecting the presence of
CC polymorphisms associated with inflammatory bowel diseases such as
CC ulcerative colitis and Crohn's disease. The methods can be used to detect
CC the presence of genetic polymorphisms associated with inflammatory bowel
CC disease and correlating their occurrence with disease states. They may be
CC used in this way for phenotypic correlations, forensics, paternity
CC testing, medicine and genetic analysis. The present sequence is a gene
CC containing a polymorphic site described in the exemplification of the
XX invention.
XX
XX
SQ Sequence 699 BP; 192 A; 142 C; 99 G; 266 T; 0 other;

Query Match 5.4%; Score 17; DB 22; Length 699;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 aacaatgaatcagtt 60
DB 372 aacaatgaatcagtt 388

RESULT 11
AAH04423
ID AAH04423 standard; cDNA; 726 BP.
XX
XX AAH04423;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA clone (5'-primer) SEQ ID NO:1258.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX

```

PN  EPI074617-A2.
XX
XX  07-FEB-2001.
XX
XX  28-JUL-2000; 2000EP-0116126.
XX
XX  29-JUL-1999; 99JP-0248036.
XX  27-AUG-1999; 99JP-0300253.
XX  11-JAN-2000; 2000JP-0118776.
XX  02-MAY-2000; 2000JP-0183767.
XX  09-JUN-2000; 2000JP-0241899.
XX
XX  (HELI-) HELIX RES INST.
XX
XX  Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI  Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX  WPI; 2001-318749/34.
XX
XX  Primer sets for synthesizing polynucleotides, particularly the 5602
PT  full-length cDNAs defined in the specification, and for the detection
PT  and/or diagnosis of the abnormality of the proteins encoded by the
PT  full-length cDNAs -
XX
XX  Claim 1; SEQ ID 1258; 2537bp + CD ROM; English.
XX
XX  The present invention describes primer sets for synthesizing 5602
XX  full-length cDNAs defined in the specification. Where a primer set
XX  comprises: (a) an oligo-dt primer and an oligonucleotide complementary
XX  to the complementary strand of a polynucleotide which comprises one of
XX  the 5602 nucleotide sequences defined in the specification, where the
XX  oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX  of an oligonucleotide comprising a sequence complementary to the
XX  complementary strand of a polynucleotide which comprises a 5'-end
XX  sequence and an oligonucleotide comprising a sequence complementary to a
XX  polynucleotide which comprises a 3'-end sequence, where the
XX  oligonucleotide comprises at least 15 nucleotides, and the combination of
XX  the 5'-end sequence/3'-end sequence is selected from those defined in
XX  the specification. The primer sets can be used in antisense therapy and
XX  in gene therapy. The primers are useful for synthesizing polynucleotides,
XX  particularly full-length cDNAs. The primers are also useful for the
XX  detection and/or diagnosis of the abnormality of the proteins encoded by
XX  the full-length cDNAs. The primers allow obtaining of the full-length
XX  cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX  AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX  AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX  represent oligonucleotides, all of which are used in the exemplification
XX  of the present invention.
XX
XX  Sequence 726 BP; 252 A; 109 C; 145 G; 217 T; 3 other.
XX
XX  Query Match          5.4%; Score 17; DB 22; Length 726;
XX  Best Local Similarity 100.0%; Pred. No. 70;
XX  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY  175 caaatlttgaataaga 191
XX  DB  173 caaatlttgaataaga 189
XX
XX  RESULT 12
XX  AAZ36779
XX  ID  AAZ36779 standard; DNA; 842 BP.
XX
XX  AC  AAZ36779;
XX
XX  DT  13-MAR-2000 (first entry)
XX
XX  DE  Nucleotide sequence of the human AHCP gene clone 3.
XX
XX  KW  Human; AHCP gene; autosomal highly conserved protein; schizophrenia;
KW  neurological disease; genetic predisposition; chromosome 6p23; D6S274;

```

```

KW  D6S285; psychological disease; gene therapy; ss.
XX
XX  Homo sapiens.
XX
XX  WO9957316-A1.
XX
XX  11-NOV-1999.
XX
XX  PD  30-APR-1999; 99WO-IB00846.
XX
XX  PE  30-APR-1999; 99WO-IB00846.
XX
XX  PR  30-APR-1999; 98US-0083625.
XX  31-DEC-1998; 98US-0114592.
XX
XX  PA  (INSP ) INST PASTEUR.
XX  (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX  Leroy P, Bourgeron T, McElreavey K, Fellous M, Jamain S;
XX
XX  WPI; 2000-086415/07.
XX
XX  DR  WPI; 2000-086415/07.
XX
XX  PT  New gene encoding autosomal high conserved protein used to diagnose a
PT  genetic predisposition to schizophrenia -
XX
XX  Example 1; Page 66-67; 76pp; English.
XX
XX  PS  AAZ36777-84 represent overlapping clones of the complete human AHCP
XX  CC  (autosomal highly conserved protein) gene. The AHCP gene is linked
XX  CC  to a genetic predisposition to schizophrenia. The gene is located on
XX  CC  chromosome 6p23, between markers D6S274 and D6S285. Several
XX  CC  polymorphisms are found in the AHCP gene. Oligonucleotide probes
XX  CC  derived from the AHCP sequences can be used to screen for patients
XX  CC  having a genetic predisposition for a neurological or psychological
XX  CC  disease, especially schizophrenia. The invention is used to diagnose
XX  CC  a genetic predisposition to schizophrenia, and to treat the disorder
XX  CC  by gene therapy. The invention provides a treatment that is specific to
XX  CC  schizophrenic disorders, without the risk of significant side effects.
XX
XX  SQ  Sequence 842 BP; 258 A; 135 C; 158 G; 291 T; 0 other.
XX
XX  Query Match          5.4%; Score 17; DB 21; Length 842;
XX  Best Local Similarity 100.0%; Pred. No. 70;
XX  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY  15 gattttaagaataact 31
XX  DB  283 gattttaagaataact 299
XX
XX  RESULT 13
XX  AAH94373/C
XX  ID  AAH94373 standard; cDNA; 1239 BP.
XX
XX  AC  AAH94373;
XX
XX  DT  05-OCT-2001 (first entry)
XX
XX  DE  Human foetal cDNA, SEQ ID NO: 902.
XX
XX  KW  Human; foetal protein; cytosolic; immunosuppressive; immunostimulant;
KW  neurotrophic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW  gene therapy; antisense therapy; cancer; immune disorder;
KW  growth disorder; osteoporosis; thrombolytic disorder;
KW  nervous system disorder; inflammation; expressed sequence tag; EST; ss.
XX
XX  OS  Homo sapiens.
XX
XX  PN  WO200155339-A2.
XX
XX  PD  02-AUG-2001.
XX
XX  PF  25-JAN-2001; 2001WO-US02723.
XX

```

PR 25-JAN-2000; 2000US-0491404.
PR 15-SEP-2000; 2000US-0663870.
PR 06-NOV-2000; 2000US-0707351.
XX
PA (HSE-) HSEQ INC.
XX
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
PI Liu C, Asundi V, Zhou P, Werhman T;
XX
DR WPI: 2001-465571/50.
DR P-PSDB: AAM06698.
XX
PT Novel fetal proteins useful for the treatment and diagnosis of diseases
PT associated with dysfunction of the protein e.g. cancers, immune
PT disorders, growth disorders, thrombolytic disorders, nervous system
PT disorders and inflammation -
XX
PS Claim 1; Page 526-527; 715pp; English.
XX
CC The invention relates to novel foetal polypeptides encoded by
CC polynucleotides comprising one of 477 sequences fully defined in the
CC specification. The foetal polynucleotides and polypeptides are
CC useful in the treatment and diagnosis of diseases such as cancers,
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC disorders, nervous system disorders and inflammation. The present
CC sequence is a full length cDNA which was assembled using expressed
CC sequence tags (ESTs) found to be expressed in human foetal tissue
CC cDNA libraries as seeds.
XX
SQ Sequence 1239 BP; 370 A; 187 C; 246 G; 436 T; 0 other;

Query Match 5.4%; Score 17; DB 22; Length 1239;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 34 tgaagaatagaacaat 50
|||
Db 1239 TCGAATAATGACAAAT 1223

RESULT 14
ABA45382
ID ABA45382 standard; DNA: 1959 BP.
XX
AC ABA45382;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #4077.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX

XX
DR WPI: 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 1; SEQ ID NO 4077; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WPI at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1959 BP; 521 A; 343 C; 564 G; 531 T; 0 other;

Query Match 5.4%; Score 17; DB 22; Length 1959;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 124 agtgggtgccttgt 140
|||
Db 270 agtgggtgccttgt 286

RESULT 15
ABA55872
ID ABA55872 standard; DNA: 1959 BP.
XX
AC ABA55872;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #4177.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-483447/52.
XX

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 1: SEQ ID NO 4177; 639bp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1959 BP; 521 A; 343 C; 564 G; 531 T; 0 other;

Query Match 5.48; Score 17; DB 22; Length 1959;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 124 agtgggtgtgcttctgt 140
 |||||
 Db 270 agtgggtgtgcttctgt 286

Search completed: September 16, 2002, 06:31:41
 Job time: 7649 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:24:15 ; Search time 90.91 Seconds

(without alignments)
851.112 Million cell updates/sec

Title: US-09-700-770-4

Perfect score: 315

Sequence: 1 taacacactgactcagattt.....tgtgtattacttactatgc 315

Scoring table: OLIGO_NUC

Searched: 38353 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2-6/ptodata/2/1na/5A_COMB.seq: *
2: /cgn2-6/ptodata/2/1na/5B_COMB.seq: *
3: /cgn2-6/ptodata/2/1na/6A_COMB.seq: *
4: /cgn2-6/ptodata/2/1na/6B_COMB.seq: *
5: /cgn2-6/ptodata/2/1na/PTUS_COMB.seq: *
6: /cgn2-6/ptodata/2/1na/Backfill1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 18 | 5.7 | 9620 | 4 | US-08-952-127-11 |
| 2 | 17 | 5.4 | 4146 | 1 | US-08-261-822A-15 |
| 3 | 17 | 5.4 | 4146 | 5 | PCT-US95-07744A-15 |
| 4 | 16 | 5.1 | 715 | 4 | US-08-991-789A-264 |
| 5 | 16 | 5.1 | 715 | 4 | US-09-062-451-264 |
| 6 | 16 | 5.1 | 1161 | 2 | US-08-616-857-1 |
| 7 | 16 | 5.1 | 2831 | 2 | US-09-906-713-1 |
| 8 | 16 | 5.1 | 6688 | 4 | US-09-381-862-5 |
| 9 | 16 | 5.1 | 9100 | 2 | US-08-743-637B-27 |
| 10 | 16 | 5.1 | 9100 | 3 | US-08-526-840B-27 |
| 11 | 15 | 4.8 | 23 | 1 | US-07-741-940-64 |
| 12 | 15 | 4.8 | 23 | 1 | US-08-289-348A-64 |
| 13 | 15 | 4.8 | 23 | 1 | US-08-452-654-64 |
| 14 | 15 | 4.8 | 23 | 1 | US-08-452-655B-64 |
| 15 | 15 | 4.8 | 23 | 3 | US-08-450-582-64 |
| 16 | 15 | 4.8 | 51 | 2 | US-08-350-260A-265 |
| 17 | 15 | 4.8 | 54 | 2 | US-08-350-260A-266 |
| 18 | 15 | 4.8 | 57 | 2 | US-08-350-260A-267 |
| 19 | 15 | 4.8 | 381 | 2 | US-08-844-153-3 |
| 20 | 15 | 4.8 | 486 | 1 | US-08-828-511-1 |
| 21 | 15 | 4.8 | 666 | 3 | US-09-081-320-15 |
| 22 | 15 | 4.8 | 714 | 4 | US-08-896-164-19 |
| 23 | 15 | 4.8 | 945 | 1 | US-08-253-155A-10 |
| 24 | 15 | 4.8 | 1006 | 3 | US-08-924-747-15 |
| 25 | 15 | 4.8 | 1006 | 4 | US-09-247-373B-15 |
| 26 | 15 | 4.8 | 1006 | 4 | US-09-296-715-15 |
| 27 | 15 | 4.8 | 1443 | 2 | US-08-962-203-1 |

| | | | | | |
|----|----|-----|------|---|--------------------|
| 28 | 15 | 4.8 | 1443 | 4 | US-09-282-125A-1 |
| 29 | 15 | 4.8 | 1443 | 4 | US-09-273-142-1 |
| 30 | 15 | 4.8 | 1569 | 2 | US-08-743-637B-176 |
| 31 | 15 | 4.8 | 1569 | 3 | US-08-526-840B-176 |
| 32 | 15 | 4.8 | 1734 | 1 | US-08-551-437-4 |
| 33 | 15 | 4.8 | 1734 | 3 | US-09-004-225-4 |
| 34 | 15 | 4.8 | 1734 | 4 | US-09-084-346-4 |
| 35 | 15 | 4.8 | 1734 | 4 | US-09-104-704-4 |
| 36 | 15 | 4.8 | 1886 | 2 | US-08-659-939-1 |
| 37 | 15 | 4.8 | 1886 | 2 | US-08-850-041-1 |
| 38 | 15 | 4.8 | 1886 | 2 | US-08-467-963C-5 |
| 39 | 15 | 4.8 | 1886 | 2 | US-08-838-189D-5 |
| 40 | 15 | 4.8 | 1886 | 3 | US-08-852-344D-5 |
| 41 | 15 | 4.8 | 1886 | 3 | US-08-896-500-1 |
| 42 | 15 | 4.8 | 1886 | 3 | US-08-476-397-1 |
| 43 | 15 | 4.8 | 1886 | 3 | US-08-973-720-1 |
| 44 | 15 | 4.8 | 1886 | 3 | US-08-344-639E-5 |
| 45 | 15 | 4.8 | 1886 | 3 | US-09-262-927-1 |

ALIGNMENTS

RESULT 1
US-08-952-127-11
; Sequence 11, Application US/08952127
; Patent No. 6211336
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kohn & Associates
; STREET: 30500 No. 6211336thwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,127
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,995
; REFERENCE/DOCKET NUMBER: 2290.00029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 810-539-5050
; TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO. 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: Chromosome 9, Band 9C
; US-08-952-127-11
Query Match 5.7%; Score 18; DB 4; Length 9620;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 tctcttagagatcctacag 97
|||||
Db 6071 TCTCTTAGAGATCTACAG 6088

RESULT 2

US-08-261-822A-15
Sequence 15, Application US/08261822A
Patent No. 5650553
GENERAL INFORMATION:
APPLICANT: Ecker, Joseph R. et al.
TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene
TITLE OF INVENTION: and Pathogens
NUMBER OF SEQUENCES: 82
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5650553-15
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,822A
FILING DATE: 17-JUN-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Beardell, Lori Y.
REGISTRATION NUMBER: 34,293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3439
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 4146 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-261-822A-15

Query Match 5.4%; Score 17; DB 1; Length 4146;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 176 aaatttgaataatagat 192
|||||
Db 1284 AAATTTGAAATAGAT 1300

RESULT 3

PCT-US95-07744A-15
Sequence 15, Application PC/TUS9507744A
GENERAL INFORMATION:
APPLICANT: Trustees of The University of Pennsylvania
TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene
TITLE OF INVENTION: and Pathogens
NUMBER OF SEQUENCES: 82
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07744A
FILING DATE: 15-JUNE-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,822
FILING DATE: June 17, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Beardell, Lori Y.
REGISTRATION NUMBER: 34,293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3439
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 4146 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-07744A-15

Query Match 5.4%; Score 17; DB 5; Length 4146;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 176 aaatttgaataatagat 192
|||||
Db 1284 AAATTTGAAATAGAT 1300

RESULT 4

US-08-991-789A-264/C
Sequence 264, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 264:
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 264:
US-08-991-789A-264

Query Match 5.1%; Score 16; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 ttgaaatagattgt 195
|||||
Db 289 TTGAAATAGATTGT 274

RESULT 5
US-09-062-451-264/C

Sequence 264, Application US/09062451
Patent No. 6344350
GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 264:
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-264

Query Match 5.1%; Score 16; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 ttgaaatagattgt 195
|||||
Db 289 TTGAAATAGATTGT 274

RESULT 6
US-08-616-857-1/C

Sequence 1, Application US/08616857
Patent No. 5905146
GENERAL INFORMATION:
APPLICANT: Lecka-Czerwik, Dr. Beata
TITLE OF INVENTION: NOVEL DNA BINDING PROTEIN S1-3
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/616,857
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John F.
REGISTRATION NUMBER: 29,066
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2..674
US-08-616-857-1

Query Match 5.1%; Score 16; DB 2; Length 1161;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 ttgaaatagattgt 195
|||||
Db 816 TTGAAATAGATTGT 801

RESULT 7
US-08-906-713-1
Sequence 1, Application US/08906713
Patent No. 5965704
GENERAL INFORMATION:
APPLICANT: Lok, Si
APPLICANT: Adams, Robyn L.
APPLICANT: Jelmsberg, Anna C.
APPLICANT: Whitmore, Theodore E.
APPLICANT: Farrah, Theresa M.
TITLE OF INVENTION: MAMMALIAN ZCYTOR11
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics
STREET: 1201 Eastlake Ave East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible

OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,713
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lunn, Paul G
REGISTRATION NUMBER: 32,743
REFERENCE/DOCKET NUMBER: 97-52
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6627
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2831 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 34...1755
OTHER INFORMATION:
US-08-906-713-1

Query Match 5.1%; Score 16; DB 2; Length 2831;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 aacaaatgaatcact 59
|||||
Db 2631 AACAAATGAAATCAGT 2646

RESULT 8
US-09-381-862-5/c
Sequence 5, Application US/09381862
Patent No. 6245906
GENERAL INFORMATION:
APPLICANT: Ueyama, Hiroshi
APPLICANT: Abe, Kanako
APPLICANT: Keshi, Hiroyuki
APPLICANT: Matsubisa, Akio
TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS
TITLE OF INVENTION: CAUSED BY STREPTOCOCCUS PYOGENES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
City: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/381,862
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1997-71077
FILING DATE: 25-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP98/01288

FILING DATE: 23-MAR-1998
ATTORNEY/AGENT INFORMATION:
NAME: Cawley, Jr., Thomas A.
REGISTRATION NUMBER: 40,944
REFERENCE/DOCKET NUMBER: 19036/36274
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 6688 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
STRAIN: Clinical Isolate SP-26-46
US-09-381-862-5

Query Match 5.1%; Score 16; DB 4; Length 6688;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 aaataactttgagaa 39
|||||
Db 5369 AAATAACTTTGAGAA 5354

RESULT 9
US-08-743-637B-27
Sequence 27, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414)277-5591
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 9100 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Haemophilus influenzae
US-08-743-637B-27

Query Match 5.1%; Score 16; DB 2; Length 9100;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 234 ggaataacacattcc 249
|||||
Db 3581 GGAATACACATTCC 3596

RESULT 10
US-08-526-840B-27
Sequence 27, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: OUALES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 9100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Haemophilus influenzae
US-08-526-840B-27

Query Match 5.1%; Score 16; DB 3; Length 9100;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 234 ggaataacacattcc 249
|||||
Db 3581 GGAATACACATTCC 3596

RESULT 11
US-07-741-940-64
Sequence 64, Application US/07741940
Patent No. 5352775
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-07-741-940-64

Query Match 4.8%; Score 15; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 199 acaataactgagct 213
|||||
Db 1 ACAATAACTGAGCT 15

RESULT 12
US-08-289-548A-64
Sequence 64, Application US/08289548A
Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagau, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ. ID NO.: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-289-548A-64

Query Match 4.8%; Score 15; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 acaataactgagat 213
|||||
Db 1 ACAATTAAGTGAGT 15

RESULT 13
US-08-452-654-64
Sequence 64, Application US/08452654
Patent No. 5691454
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598

NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagau, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ. ID NO.: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-452-654-64

Query Match 4.8%; Score 15; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 acaataactgagat 213
|||||
Db 1 ACAATTAAGTGAGT 15

RESULT 14
US-08-452-655B-64
Sequence 64, Application US/08452655B
Patent No. 5783666
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-452-655B-64

Query Match 4.8%; Score 15; DB 1; Length 23;
Best local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 199 acataaactgagct 213
|||||
Db 1 ACAATAAAGTGAGCT 15

RESULT 15
US-08-450-582-64
Sequence 64, Application US/08450582
Patent No. 6114124
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THIVIERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,582
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/452,655
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-450-582-64

Query Match 4.8%; Score 15; DB 3; Length 23;
Best local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 199 acataaactgagct 213
|||||
Db 1 ACAATAAAGTGAGCT 15

Search completed: September 16, 2002, 06:24:21
Job time: 10634 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 05:18:12 ; Search time 3017.73 Seconds
(without alignments)
1408.854 Million cell updates/sec

Title: US-09-700-770-4
Perfect score: 315
Sequence: 1 taacactgactcagattt.....tggtatatactactatgc 315

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 674847542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_estlhm:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 264 | 83.8 | 563 | 10 | BF335657 RCI-CT046 |
| 2 | 81 | 25.7 | 513 | 10 | N49446 yv22a12.r1 |
| 3 | 42 | 13.3 | 323 | 9 | AM753759 RC3-CT028 |
| 4 | 21 | 6.7 | 573 | 12 | BH11628 RCI-24-2 |
| 5 | 21 | 6.7 | 685 | 12 | AO780202 HS-3138.B |
| 6 | 20 | 6.3 | 578 | 12 | A2319774 IM0039F04 |
| 7 | 19 | 6.0 | 205 | 9 | AV163511 |
| 8 | 19 | 6.0 | 258 | 10 | BG999039 |
| 9 | 19 | 6.0 | 300 | 10 | BG99033 PMO-HT091 |
| 10 | 19 | 6.0 | 364 | 10 | C56866 C56866 Yuj1 |
| 11 | 19 | 6.0 | 376 | 10 | BM431586 IDuc24D11 |
| 12 | 19 | 6.0 | 416 | 10 | BF711377 MI-P-A1-a |
| 13 | 19 | 6.0 | 510 | 10 | AO669043 HS-5414.A |
| 14 | 19 | 6.0 | 585 | 12 | B1075543 B1075543 |
| 15 | 19 | 6.0 | 651 | 9 | BH524918 BCGXTC4TF |
| 16 | 19 | 6.0 | 680 | 12 | BB656024 BB656024 |
| 17 | 19 | 6.0 | 755 | 12 | A2521650 RCI-11-2 |
| | | | | | BH022291 GH_MBB000 |

| | | | | | | |
|----|----|-----|-----|----|----------|--------------------|
| 18 | 19 | 6.0 | 788 | 12 | BH250476 | BH250476 BCGAC49TF |
| 19 | 19 | 6.0 | 801 | 12 | BH544046 | BH544046 BOHGO03TF |
| 20 | 19 | 6.0 | 831 | 12 | AZ208098 | AZ208098 SP_0136.B |
| 21 | 18 | 5.7 | 212 | 12 | AZ603387 | AZ603387 INQ422P17 |
| 22 | 18 | 5.7 | 266 | 9 | AV074635 | AV074635 AV074635 |
| 23 | 18 | 5.7 | 288 | 9 | AV072249 | AV072249 AV072249 |
| 24 | 18 | 5.7 | 309 | 10 | B1080659 | B1080659 B1080659 |
| 25 | 18 | 5.7 | 335 | 9 | AA834642 | AA834642 od98C03.s |
| 26 | 18 | 5.7 | 364 | 9 | AA206457 | AA206457 zq58a12.s |
| 27 | 18 | 5.7 | 380 | 9 | AA495343 | AA495343 fa02a01.r |
| 28 | 18 | 5.7 | 393 | 10 | B1467110 | B1467110 1c52912.x |
| 29 | 18 | 5.7 | 393 | 10 | B1790038 | B1790038 1c44b08.x |
| 30 | 18 | 5.7 | 394 | 10 | B1863517 | B1863517 kx45b09.y |
| 31 | 18 | 5.7 | 423 | 10 | B1789842 | B1789842 1c44b08.y |
| 32 | 18 | 5.7 | 436 | 12 | AZ812831 | AZ812831 2M0079L17 |
| 33 | 18 | 5.7 | 450 | 10 | B1441375 | B1441375 1c52912.y |
| 34 | 18 | 5.7 | 455 | 10 | BE919766 | BE919766 EST423535 |
| 35 | 18 | 5.7 | 457 | 10 | B1349672 | B1349672 dae62c01. |
| 36 | 18 | 5.7 | 457 | 10 | T54305 | T54305 ya91b03.s3 |
| 37 | 18 | 5.7 | 464 | 10 | BF152600 | BF152600 uz35e04.y |
| 38 | 18 | 5.7 | 471 | 10 | B1043094 | B1043094 B1043094 |
| 39 | 18 | 5.7 | 479 | 10 | BM257055 | BM257055 520551.MA |
| 40 | 18 | 5.7 | 479 | 12 | AO692875 | AO692875 HS-5414.B |
| 41 | 18 | 5.7 | 486 | 10 | BE677407 | BE677407 7d83g11.x |
| 42 | 18 | 5.7 | 493 | 10 | BF707366 | BF707366 282927.MA |
| 43 | 18 | 5.7 | 500 | 9 | AU088111 | AU088111 AU088111 |
| 44 | 18 | 5.7 | 500 | 10 | BF440800 | BF440800 256827.MA |
| 45 | 18 | 5.7 | 501 | 10 | BF440803 | BF440803 256833.MA |

ALIGNMENTS

RESULT 1
BF335657
LOCUS RCI-CT0469-160800-011-e04 CT0469 Homo sapiens CDNA, mRNA sequence.
DEFINITION BF335657
ACCESSION BF335657.1 GI:11306405
VERSION
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 563)
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

TEL: +55-11-2704922
FAX: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1ct2-RC1-CT0469-160800-011-e04&t3=2000-08-16&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 21
High quality sequence stop: 559.

Location/Qualifiers
1. 563

FEATURES
source
/organism="Homo sapiens"


```

source
1. 323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="C0283"
/dev_stage="Adult"
/note="Organ: Colon; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      84 a      59 c      61 g      119 t
ORIGIN

Query Match      13.3%; Score 42; DB 9; Length 323;
Best Local Similarity 100.0%; Pred. No. 1e-10;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      18      tttaagaataactttgaagaatgaacaaatgaatacagt 59
Db      64      TTTAAGAATATACCTTTTGAGAAATGAGACAAATGAATCAGT 23

RESULT      4
BH11628
LOCUS      BH11628
DEFINITION  BH11628 573 bp DNA linear GSS 19-JUL-2001
            RPCI-24-252C22.TJ RPCI-24 Mus musculus genomic clone RPCI-24-252C22
            , DNA sequence.
ACCESSION   BH11628
VERSION     BH11628.1 GI:14946957
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 573)
            Zhao, S., Niernan, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
            Tesgeye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregeorgis, E.,
            Russell, D., de Jong, P. and Fraser, C.M.
            Mouse BAC End Sequences from Library RPCI-24
            Unpublished (1999)
            Other GSSs: RPCI-24-252C22.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 252 row: C column: 22
            Seq primer: SP6
            Class: BAC ends.
FEATURES
source      Location/Qualifiers
1. 573
/organism="Mus musculus"
/db_xref="taxon:9606"
/clone_lib="C57BL/6J"
/clone="RPCI-24-252C22"
/clone_id="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1.
The library was cloned in the pTARBAC1 cloning vector at the
BamH1 sites using KmoI partially digested male C57BL/6J
DNA."

```

```

BASE COUNT      190 a      103 c      124 g      156 t
ORIGIN

Query Match      6.7%; Score 21; DB 12; Length 573;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      14      agatttgaataactttt 34
Db      133      AGATTTTAAGAAATACCTTTT 153

RESULT      5
AO780202
LOCUS      AO780202
DEFINITION  HS_3138_B1_H06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3138 COL=11 Row=P, DNA sequence.
ACCESSION   AO780202
VERSION     AO780202.1 GI:5683162
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 685)
            Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
            Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
            Hood, L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
            Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
            99380589
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web server: http://www.hsc.washington.edu
            Plate: 3138 row: P column: 11
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 685.
FEATURES
source      Location/Qualifiers
1. 685
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3138 COL=11 Row=P"
/clone_id="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC clones in
E-Coli DH10B"
E-Coli DH10B"
BASE COUNT      284 a      132 c      94 g      174 t      1 others
ORIGIN

Query Match      6.7%; Score 21; DB 12; Length 685;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      112      acatacaaatgcagtcgtcgtt 132
Db      26      ACATCAAAAGTCAGTTGTGTT 46

RESULT      6
AZ319774
LOCUS      AZ319774
DEFINITION  IM0039F04R Mouse 10kb plasmid UUC1M library Mus musculus genomic
            clone UUC1M0039F04 R, DNA sequence.

```

ACCESSION A2319774
 VERSION A2319774.1 GI:10370884
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 578)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0039 row: F column: 04
 Seq primer: CACACGGAACACGCTATGACC
 Class: Plasmid ends
 High quality sequence stop: 578.
 Location/Qualifiers
 1..578
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U06C1M0039F04"
 /clone_lib="Mouse 10kb plasmid U06C1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male); was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114/gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 188 a 106 c 149 g 135 t
 ORIGIN

Query Match 6.3%; Score 20; DB 12; Length 578;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 229 tagaaggaataacacattc 248
 ||||||||||||||||
 Db 63 TAGAAGGAATTCACACATTC 82

RESULT 7
 LOCUS AV163511
 DEFINITION AV163511 Mus musculus head C57BL/6J 13-day embryo Mus musculus cDNA

ACCESSION clone 3110011B18, mRNA sequence.
 VERSION AV163511
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 205)
 AUTHORS Carrinot,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tateo,M., Tomaru,Y., Tomimaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
 TITLE RIKEN Mouse ESTs
 JOURNAL Unpublished (1999)
 COMMENT Contact: Chie Owa
 Genome Science Laboratory
 RIKEN
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-9145
 Fax: 81-298-36-9098
 Email: genome-res@rtc.riken.go.jp
 Thermostabilization and thermooactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
 Location/Qualifiers
 1..205
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="3110011B18"
 /clone_lib="Mus musculus head C57BL/6J 13-day embryo"
 /sex="mixed"
 /tissue_type="head"
 /dev_stage="13-day embryo"
 /base_count="31 c 37 g 73 t
 BASE COUNT 64 a 31 c 37 g 73 t
 ORIGIN

Query Match 6.0%; Score 19; DB 9; Length 205;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 59 ttctccacacacttaagta 77
 ||||||||||||||||
 Db 17 TTTCTCCACACACTTAAGTA 35

RESULT 8
 LOCUS BG999039/c 258 bp mRNA linear EST 13-JUN-2001
 DEFINITION PMO-HR0913-250401-015-C04 HT0913 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG999039
 VERSION BG999039.1 GI:14403111
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 258)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W., Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

TITLE Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 2020263
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM0&t2=PM0-HT0913-
250401-015-C04&t3=2001-04-25&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 22
High quality sequence stop: 258.

FEATURES
source
1. 258
location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0913"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 85 a 53 c 55 g 65 t
ORIGIN

Query Match 6.0%; Score 19; DB 10; Length 258;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 cagatttaagaataact 31
|||||
Db 226 CAGATTTAAGAAATTA 208

RESULT 9
C56866 300 bp mRNA linear EST 22-SEP-1997
LOCUS C56866 yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA

DEFINITION Clone yk278b6 3', mRNA sequence.
ACCESSION C56866
VERSION C56866.1 GI:2414897

KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE 1 (bases 1 to 300)
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides
; Rhabditidae; Pelodirinae; Caenorhabditis.
Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
M., Miyata,A. and Nishigaki,A.
Expression map of the C.elegans genome
Unpublished (1996)

TITLE Contact: Yuji Kohara
JOURNAL Genome Biology Lab.
COMMENT National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1. 300
/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone_lib="yk278b6"
/clone_lib="yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 138 a 29 c 59 g 70 t 4 others
ORIGIN

Query Match 6.0%; Score 19; DB 10; Length 300;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 168 atgactacaatttgaag 186
|||||
Db 71 ATGACTACAAATTTGAAA 89

RESULT 10
BM431586 364 bp mRNA linear EST 31-JAN-2002
LOCUS BM431586
DEFINITION 1duo24D11.ab1 Bos taurus Duodenum #1 library Bos taurus cDNA, mRNA
sequence.

ACCESSION BM431586
VERSION BM431586.1 GI:18453308
KEYWORDS EST.
SOURCE

ORGANISM Bos taurus
cow.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 364)
Hansen,C., Fu,A., Weng,Y., Li,C., Okine,E., Sensen,C.W., Gordon
P.M.K. and Moore,S.S.
Gene Expression Profiling of the Bovine Gastrointestinal Tract
Unpublished (2002)

TITLE

JOURNAL Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265

COMMENT Email: smoores@afns.ualberta.ca
Insert Length: 364 Std Error: 0.00
POLYA=Yes.

FEATURES
source
1. 364
location/Qualifiers

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bos taurus Duodenum #1 library"
/tissue_type="Smooth muscle"
/cell_type="Simple columnar epithelial"
/dev_stage="Young adult"
/lab_host="X11-BlueMF's strain"
/note="Organ: Intestine/Duodenum; Vector: Uni-ZAPvXR;
Site_1: EcoRI; Site_2: Xho I"
BASE COUNT 123 a 69 c 62 g 110 t
ORIGIN

Query Match 6.0%; Score 19; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 169 ttactacaatttgaag 187
|||||
Db 343 TGACTACAAATTTGAAA 361

RESULT 11
BF711377/c

LOCUS BF711377 376 bp mRNA linear EST 02-JAN-2001
 DEFINITION MI-P-AI-acm-g-11-1-UM.s1 MI-P-AI Sus scrofa cDNA clone
 ACCESSION MI-P-AI-acm-g-11-1-UM 3', mRNA sequence.
 VERSION BF711377
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 REFERENCE 1 (bases 1 to 376)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tugayle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: ctkugley@iastate.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 non-normalized anterior pituitary at estrus day 0 library cDNA
 library preparation: RJ Woods, JA Green, RS Prather, S142 Animal
 Science Research Center, Department of Animal Science, University
 of Missouri-Columbia, 65211 Clone distribution: clones will be
 available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source Location/Qualifiers
 1..376
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="MI-P-AI-acm-g-11-1-UM"
 /clone_lib="MI-P-AI"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p77T3D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: EcoRI; The MI-P-AI
 library is derived from anterior pituitary at estrus day
 0. For a detailed description of the library from which
 this clone was derived, please visit our web site at
 http://pigest.genome.iastate.edu/."
 TAG_LIB="MI-P-AI"
 TAG_TISSUE="anterior pituitary at estrus day 0"
 TAG_SEQ="TAAGGC"
 BASE COUNT 133 a 65 c 57 g 121 t
 ORIGIN

Query Match 6.0%; Score 19; DB 10; Length 376;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 gtatgtatgactacaatt 180
 ||||||||||||||||
 Db 280 GTAGTATGACTACAAATT 262

RESULT 12
 LOCUS AG669043 416 bp DNA linear GSS 24-JUN-1999
 DEFINITION HS_5414_A2_A09_SP6E RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate=990 Col=18 Row=A, DNA sequence.
 ACCESSION AG669043
 VERSION AG669043.1 GI:5201789
 KEYWORDS GSS.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 416)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
 or from Research Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 990 row: A column: 18
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 416.

FEATURES
 source Location/Qualifiers
 1..416
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=990 Col=18 Row=A"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACe3.6 vector at EcoRI sites"
 BASE COUNT 187 a 64 c 64 g 101 t
 ORIGIN

Query Match 6.0%; Score 19; DB 12; Length 416;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 tcagataattgtcatgt 298
 ||||||||||||||||
 Db 125 TCAGTATATTGTCTATGT 107

RESULT 13
 LOCUS BJ075543 510 bp mRNA linear EST 11-DEC-2001
 DEFINITION BJ075543 N1BB Mochli normalized Xenopus tailbud library Xenopus
 laevis cDNA clone XL057004 5', mRNA sequence.
 ACCESSION BJ075543
 VERSION BJ075543.1 GI:17520459
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 510)
 AUTHORS Kitayama,A., Terasaka,C., Mochli,M., Ueno,N., Shin-I,T. and Kohara
 Y.
 TITLE Expressed genes in X. laevis embryo
 JOURNAL Unpublished (2001)
 COMMENT Contact: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856
Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

1. .510

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="XL057004"

/clone_lib="NIBB Mochii normalized Xenopus tailbud library"

/tissue_type="whole embryo"

/dev_stage="stage 25"

BASE COUNT 156 a 107 c 110 g 137 t

ORIGIN

Query Match 6.0%; Score 19; DB 10; Length 510;

Best Local Similarity 100.0%; Pred. No. 56;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 tagaaggaataacacact 247

Db 355 TAGAAGGAATACACACTT 337

RESULT 14

BH524918

LOCUS BH524918 585 bp DNA linear GSS 13-DEC-2001

DEFINITION BOGXT64TF BOG Brassica oleracea genomic clone BOGXT64, DNA

ACCESSION BH524918

VERSION BH524918

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 585)

Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other GSS: BOGXT64TR

COMMENT

Contact: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1. .585

/organism="Brassica oleracea"

/strain="T01000DH3"

/db_xref="taxon:3712"

/clone="BOGXT64"

/clone_lib="BOGX"

/note="Vector: PHOSI, Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into PHOSI using BstXI linkers"

BASE COUNT 188 a 116 c 110 g 171 t

ORIGIN

Query Match 6.0%; Score 19; DB 12; Length 585;

Best Local Similarity 100.0%; Pred. No. 57;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 tgaatcagttctccacc 68

Db 490 TGAATCAGTTCTCCACC 472

RESULT 15

BB656024

LOCUS

DEFINITION BB656024 651 bp mRNA linear EST 26-OCT-2001

BB656024 RIKEN full-length enriched, 12 days embryo spinal ganglion

Mus musculus cDNA clone D130056E14 5', mRNA sequence.

ACCESSION BB656024

VERSION BB656024

KEYWORDS BB656024.1 GI:16489852

SOURCE EST.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 651)

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.

, Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda

, M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,

Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki

, D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,

Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,

Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp

URL: http://genome-gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh

, M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagii,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura

, S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and

Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

, Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa

, K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and

Hayashizaki,Y.

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome-gsc.riken.go.jp>) for

further details.

e mouse tissues.

Location/Qualifiers

1. .651

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="D130056E14"

/clone_lib="RIKEN full-length enriched, 12 days embryo

spinal ganglion"

/tissue_type="spinal ganglion"

/dev_stage="12 days embryo"

/lab_host="DH10B"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGCGGGCGGCAACACAGACTTTTTTTTTTTTTTT 3'], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGACGTTAATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from lambda FLC I."

| | | | | |
|------------|--------|-------|-------|-------|
| BASE COUNT | 217 a. | 103 c | 111 g | 220 t |
| ORIGIN | | | | |

```
Query Match      6.0%; Score 19; DB 9; Length 651;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 197 acacataaactgagtt 215
      |||||
Db 146 ACACAATAAAGTGGAGTTT 164
```

Search completed: September 16, 2002, 05:18:16
Job time: 8389 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:26:41 ; Search time 89 Seconds
(without alignments)
869.377 Million cell updates/sec

Title: US-09-700-770-4

Perfect score: 315
Sequence: 1 taacactgcctcagatttt.....tgtgtattacttactatgc 315

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/ptodata/2/lna/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/lna/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/lna/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/lna/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/lna/PTC05.COMB.seq: *
6: /cgn2_6/ptodata/2/lna/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------------|-------------------|
| C 1 | 31.2 | 9.9 | 9919 | US-08-880-179-1 | Sequence 1, Appl |
| C 2 | 31 | 9.8 | 747 | US-08-680-726A-87 | Sequence 87, Appl |
| C 3 | 31 | 9.8 | 747 | US-09-092-409-87 | Sequence 87, Appl |
| C 4 | 31 | 9.8 | 749 | US-08-680-726A-85 | Sequence 85, Appl |
| C 5 | 31 | 9.8 | 749 | US-08-680-726A-86 | Sequence 86, Appl |
| C 6 | 31 | 9.8 | 749 | US-09-092-409-85 | Sequence 85, Appl |
| C 7 | 31 | 9.7 | 749 | US-09-092-409-86 | Sequence 86, Appl |
| C 8 | 30.6 | 9.7 | 6082 | US-09-439-313-535 | Sequence 535, App |
| C 9 | 30.2 | 9.6 | 386 | US-08-906-769-142 | Sequence 142, App |
| C 10 | 30.2 | 9.6 | 386 | US-08-906-616-142 | Sequence 142, App |
| C 11 | 30.2 | 9.6 | 386 | US-08-639-075A-142 | Sequence 142, App |
| C 12 | 30.2 | 9.6 | 386 | US-09-012-431-142 | Sequence 142, App |
| C 13 | 30.2 | 9.6 | 386 | US-09-012-692-142 | Sequence 142, App |
| C 14 | 30.2 | 9.6 | 386 | US-08-906-613-142 | Sequence 142, App |
| C 15 | 30 | 9.5 | 3891 | US-08-480-604A-27 | Sequence 27, Appl |
| C 16 | 30 | 9.5 | 3891 | US-08-405-496A-27 | Sequence 27, Appl |
| C 17 | 30 | 9.5 | 3891 | US-08-915-136-27 | Sequence 27, Appl |
| C 18 | 29.8 | 9.5 | 6140 | US-09-439-313-536 | Sequence 536, App |
| C 19 | 29.6 | 9.4 | 665 | US-08-998-416-937 | Sequence 937, App |
| C 20 | 29.6 | 9.4 | 10207 | US-08-920-812-2 | Sequence 2, Appl |
| C 21 | 29.6 | 9.4 | 10207 | US-08-920-827-2 | Sequence 2, Appl |
| C 22 | 29.6 | 9.4 | 10207 | US-08-921-177-2 | Sequence 2, Appl |
| C 23 | 29.6 | 9.4 | 10207 | US-08-362-577C-2 | Sequence 2, Appl |
| C 24 | 29.6 | 9.4 | 10207 | US-08-920-828-2 | Sequence 2, Appl |
| C 25 | 29.4 | 9.3 | 51952 | US-08-947-823-1 | Sequence 1, Appl |
| C 26 | 29.2 | 9.3 | 758 | US-08-817-913-10 | Sequence 10, Appl |
| C 27 | 29.2 | 9.3 | 855 | US-08-817-913-11 | Sequence 11, Appl |

| | | | | | |
|------|------|-----|-------|------------------|-------------------|
| C 28 | 29.2 | 9.3 | 859 | US-08-817-913-12 | Sequence 12, Appl |
| C 29 | 29.2 | 9.3 | 1214 | US-08-817-913-13 | Sequence 13, Appl |
| C 30 | 29.2 | 9.3 | 1232 | US-08-817-913-14 | Sequence 14, Appl |
| C 31 | 29.2 | 9.3 | 1352 | US-08-817-913-15 | Sequence 15, Appl |
| C 32 | 29.2 | 9.3 | 1734 | US-08-817-913-16 | Sequence 16, Appl |
| C 33 | 29.2 | 9.3 | 1920 | US-08-817-913-17 | Sequence 17, Appl |
| C 34 | 28.6 | 9.1 | 797 | US-08-752-133-1 | Sequence 1, Appl |
| C 35 | 28.6 | 9.1 | 4307 | US-08-944-594-1 | Sequence 1, Appl |
| C 36 | 28.6 | 9.1 | 6216 | US-09-213-053-1 | Sequence 1, Appl |
| C 37 | 28.6 | 9.1 | 14855 | US-08-687-080-59 | Sequence 59, Appl |
| C 38 | 28.4 | 9.0 | 1789 | US-08-895-521-2 | Sequence 2, Appl |
| C 39 | 28.4 | 9.0 | 1789 | US-09-235-218-2 | Sequence 2, Appl |
| C 40 | 28.4 | 9.0 | 1981 | US-09-010-398-3 | Sequence 3, Appl |
| C 41 | 28.4 | 9.0 | 1981 | US-09-366-260-3 | Sequence 3, Appl |
| C 42 | 28.4 | 9.0 | 8224 | US-09-010-398-14 | Sequence 14, Appl |
| C 43 | 28.4 | 9.0 | 8224 | US-09-366-260-14 | Sequence 14, Appl |
| C 44 | 28.2 | 9.0 | 72604 | US-09-268-992-7 | Sequence 7, Appl |
| C 45 | 28 | 8.9 | 1529 | US-08-448-110-1 | Sequence 1, Appl |

ALIGNMENTS

RESULT 1
US-08-880-179-1/c
; Sequence 1, Application US/08880179
; Patent No. 6091004
; GENERAL INFORMATION:
; APPLICANT: Ryals, John
; APPLICANT: Delaney, Terry
; APPLICANT: Friedrich, Leslie
; APPLICANT: Weymann, Kristiana
; APPLICANT: Layton, Kay
; APPLICANT: Ellis, Daniel
; APPLICANT: Uknes, Scott
; APPLICANT: Jesse, Taco
; APPLICANT: Vos, Pieter
; TITLE OF INVENTION: GENE ENCODING A PROTEIN INVOLVED IN THE
; TITLE OF INVENTION: SIGNAL TRANSDUCTION CASCADE LEADING TO SYSTEMIC ACQUIRED RE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 6091004artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/880,179
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1909
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9919 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

1/4 acaaatcttgcgaatagattgltaccacaataacgagatltatgaaacacgcgtagaa 2330Y
20 AAAATAATTAAGAAGCGCTGTATAGTAAATCTTTATTATAGGAACCAACCAATCTCTA 79 DDB
234 ggaataacaatccatcccttcaagaagatcattactgcgaactcaggaataattgt 2930Y
80 CAAATTAATATACCTTCTTGTAATGATGATGACACCAATACGCTTTAAATATTCATCATATATTAA 1390DDB


```

; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 86:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/680,726
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 749 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-092-409-85

Query Match          9.8%; Score 31; DB 3; Length 749;
Best Local Similarity 51.9%; Pred. No. 3.3;
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 174 acaatttgaataagatgtcacacaataaactgagttttagaacaatcagtagaa 233
DB 730 AAATATATTATGAGCGCTGTTATAGCTAAATCTTTATTATATGGAACCAATCTCTA 671
QY 234 ggaatacacaaatccatcccttaccagagatcattactgaactcaggaataattgt 293
DB 670 CAAATTAATTAATCTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 611
QY 294 catgtgtattatccta 308
DB 610 TAACTATATTATCTA 596

RESULT 7
; Sequence 86, Application US/09092409
; Patent No. 6159478
; GENERAL INFORMATION:
; APPLICANT: Haanes, Elizabeth J.
; APPLICANT: Frank, Rexam S.
; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,409
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/680,726
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
```

```

; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 749 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-092-409-86
```

```

Query Match          9.8%; Score 31; DB 3; Length 749;
Best Local Similarity 51.9%; Pred. No. 3.3;
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
```

```

QY 174 acaatttgaataagatgtcacacaataaactgagttttagaacaatcagtagaa 233
DB 20 AAATATATTATGAGCGCTGTTATAGCTAAATCTTTATTATATGGAACCAATCTCTA 79
QY 234 ggaatacacaaatccatcccttaccagagatcattactgaactcaggaataattgt 293
DB 80 CAAATTAATTAATCTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 139
QY 294 catgtgtattatccta 308
DB 140 TAACTATATTATCTA 154
```

```

RESULT 8
; US-09-439-313-535/c
; Sequence 535, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 535
; LENGTH: 6082
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-439-313-535
```

```

Query Match          9.7%; Score 30.6; DB 4; Length 6082;
Best Local Similarity 68.9%; Pred. No. 8.3;
Matches 42; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```

QY 3 aacactgactcagattttagaataacttlttagaataagacaatgaatcagtttc 62
DB 4387 ACCTGATTTGATTTTAAACCAACTATGTGACATTTAAATAAATGAACCATTTTG 4328
QY 63 t 63
DB 4327 T 4327
```

```

RESULT 9
; US-08-906-769-142/c
```

```
; Sequence 142, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,769
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639, 075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..384
; OTHER INFORMATION:
; OTHER INFORMATION: R."
;
US-08-906-769-142

Query Match          9.6%; Score 30.2; DB 3; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
; Sequence 142, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,616
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..384
; OTHER INFORMATION:
; OTHER INFORMATION: R."
;
US-08-906-616-142

Query Match          9.6%; Score 30.2; DB 3; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
OY 139 gttagtcaccacttatctcaagtagtatagtacataaatttgaaatagatgtcac 198
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 280 GTACATTTACAGCCAAATAATTAATAGTATGAGAAACCAATCCACTTACCC 221
OY 199 acaataaactgagattatgtaaacatcagtagaagaagaatacaatccatccctta 258
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 220 AACCTGCCACGGTGCTATGTAACCATTAAGCCGTTTCGATTCCTTCATCAATTAACCTTTA 161
OY 259 cagagatcattactgtcaactcagagataattgt 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 160 CAGGTTTGATAGATTTTGAGTTAAACAAATTTT 126
```

RESULT 10
US-08-906-616-142/c

```
OY 139 gttagtcaccacttatctcaagtagtatagtacataaatttgaaatagatgtcac 198
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 280 GTACATTTACAGCCAAATAATTAATAGTATGAGAAACCAATCCACTTACCC 221
OY 199 acaataaactgagattatgtaaacatcagtagaagaagaatacaatccatccctta 258
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 220 AACCTGCCACGGTGCTATGTAACCATTAAGCCGTTTCGATTCCTTCATCAATTAACCTTTA 161
OY 259 cagagatcattactgtcaactcagagataattgt 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 160 CAGGTTTGATAGATTTTGAGTTAAACAAATTTT 126
```

RESULT 11
US-08-639-075A-142/c
; Sequence 142, Application US/08639075A
; Patent No. 6150125
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-APR-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..384
OTHER INFORMATION: /note="At pos. bp 3, change A to
R."
US-08-639-075A-142

Query Match 9.6%; Score 30.2; DB 3; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 139 gttgagtcacccttatattcaagtagtactacaattttgaataatgtagtcac 198
DB 280 GTACATTTACACCAATAAATATGATGTACGGAACATCCACTACTTATACCC 221
QY 199 acaataaactgagttatggaacatcagtagaagaaatacaatccatccctta 258
DB 220 AACCTGCCACGGTTGCTATTGAACCATTTAGCCGTTTCGATTCCTTCATTAAGCTTTA 161
QY 259 cagagatcattactctgcaactcagagataattgt 293
DB 160 CAGGTTTGATGATTGTGAGTTAAACAAATTTT 126

RESULT 12
US-09-012-431-142/c
; Sequence 142, Application US/09012431
; Patent No. 6180383
; GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
Rushlow, Keith E.
Wu Hunter, Shirley
Frank, Glenn R.

Stiegler, Gary
Gaines, Patrick J.
Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,431
FILING DATE: 23-Jan-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..384
OTHER INFORMATION: /note="At pos. bp 3, change A to
R."
US-09-012-431-142

Query Match 9.6%; Score 30.2; DB 4; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 139 gttgagtcacccttatattcaagtagtactacaattttgaataatgtagtcac 198
DB 280 GTACATTTACACCAATAAATATGATGTACGGAACATCCACTACTTATACCC 221
QY 199 acaataaactgagttatggaacatcagtagaagaaatacaatccatccctta 258
DB 220 AACCTGCCACGGTTGCTATTGAACCATTTAGCCGTTTCGATTCCTTCATTAAGCTTTA 161
QY 259 cagagatcattactctgcaactcagagataattgt 293
DB 160 CAGGTTTGATGATTGTGAGTTAAACAAATTTT 126

RESULT 13
US-09-012-692-142/c
; Sequence 142, Application US/09012692
; Patent No. 6214579
; GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
Rushlow, Keith E.
Wu Hunter, Shirley

```

APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary.
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESS: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
City: Denver
STATE: Colorado
COUNTRY: USA
Zip: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..384
OTHER INFORMATION:
OTHER INFORMATION: R."
US-08-906-613-142

Query Match 9.6%; Score 30.2; DB 4; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 139 gttagtcaccccttatattcaagtagtagtgaactacaatttgaataatgattgtac 198
Db 280 GTACATTTCACAGCCAAATTAATTAATTAATGATGTACGGAAACCAATCCACTTATCC 221
QY 199 acaataaacttgagattatgagaacatcagtagagaagagaatcacacattccactta 258
Db 220 AACCTGCACGCGTTGCGTAATTAACACCAATTAAGCGCGTTTGCATTCCTTCAATTA 161
QY 259 cagaatcaattacttgcgaactcaggaataattgt 293
Db 160 CAGGTTGATGATTTGAGTTTGAAGTTAAACCAATTTT 126

RESULT 15
US-08-480-604A-27/c
; Sequence 27, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.

```

```

1 APPLICANT: FIRCA, JOSEPH R.
2 APPLICANT: STAFFORD, DOUGLAS C.
3 TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
4 TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
5 NUMBER OF SEQUENCES: 32
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: MEDLEN & CARROLL, LLP
8 STREET: 220 MONTGOMERY STREET, SUITE 2200
9 CITY: SAN FRANCISCO
10 STATE: CALIFORNIA
11 COUNTRY: UNITED STATES OF AMERICA
12 ZIP: 94104
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: Patent Release #1.0, Version #1.30
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/480,604A
21 FILING DATE: 07-JUN-1995
22 CLASSIFICATION: 424
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: US 08/422,711
25 FILING DATE: 14-APR-1995
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 08/405,496
28 FILING DATE: 16-MAR-1995
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: US 08/329,154
31 FILING DATE: 25-OCT-1994
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: US 08/161,907
34 FILING DATE: 02-DEC-1993
35 PRIOR APPLICATION DATA:
36 APPLICATION NUMBER: US 07/985,321
37 FILING DATE: 04-DEC-1992
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: US 07/429,791
40 FILING DATE: 31-OCT-1989
41 ATTORNEY/AGENT INFORMATION:
42 NAME: INGOLIA, DIANE E.
43 REGISTRATION NUMBER: 40,027
44 REFERENCE/DOCKET NUMBER: OPND-01763
45 TELECOMMUNICATION INFORMATION:
46 TELEPHONE: (415) 705-8410
47 TELEFAX: (415) 397-8338
48 INFORMATION FOR SEQ ID NO: 27:
49 SEQUENCE CHARACTERISTICS:
50 LENGTH: 3891 base pairs
51 TYPE: nucleic acid
52 STRANDEDNESS: double
53 TOPOLOGY: linear
54 MOLECULE TYPE: DNA (genomic)
55 FEATURE:
56 NAME/KEY: CDS
57 LOCATION: 1...3888
58
59 US-08-480-604A-27

```

Db 2572 CTACGTATTGGAAAGCTGAAGAAGTATCTGTACTAAGTGTAT 2527

Search completed: September 16, 2002, 03:26:45
Job time: 9328 sec

| | | | | |
|--------------------------|-------|----------------|-----------|--------------|
| Query Match | 9.5% | Score 30; | DB 1; | Length 3891; |
| Best Local Similarity | 48.8% | Pred. No. 11; | | |
| Matches 81; Conservative | 0; | Mismatches 85; | Indels 0; | Gaps 0; |

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 01:42:43 : Search time 3013.46 Seconds
(without alignments)
1410.850 Million cell updates/sec

Title: US-09-700-770-4
Perfect score: 315
Sequence: 1 taacactgactcagattt.....tggtattactactatgc 315

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_hic2:*
13: em_gss_hum:*
14: em_gss_hiv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------------------|
| 1 | 313.4 | 99.5 | 563 | 10 | BF335657 RCI-CT046 |
| 2 | 132.4 | 42.0 | 513 | 10 | MA9446 yw22a12.r1 |
| 3 | 70.8 | 22.5 | 323 | 9 | AW753759 RCI-CT028 |
| 4 | 41.8 | 13.3 | 844 | 12 | CNS000FR AL069961 Drosophila |
| 5 | 40 | 12.7 | 952 | 12 | CNS001016 AL098388 Drosophila |
| 6 | 40 | 12.7 | 1101 | 12 | CNS000DT7 AL075293 Drosophila |
| 7 | 39.6 | 12.6 | 652 | 10 | BF295719 BF295719 027bfe02 |
| 8 | 39.2 | 12.4 | 1204 | 12 | CNS016E2 AL106628 Drosophila |
| 9 | 39 | 12.4 | 634 | 9 | AL639578 AL639578 |
| 10 | 39 | 12.4 | 718 | 10 | BF572670 BF572670 |
| 11 | 38.8 | 12.3 | 463 | 10 | BI360054 BI360054 384881 MA |
| 12 | 38.8 | 12.3 | 650 | 10 | BE291792 BE291792 |
| 13 | 38.4 | 12.2 | 551 | 9 | AW746159 AW746159 WSL_39.DO |
| 14 | 37.8 | 11.9 | 1225 | 12 | CNS0161D AL106171 Drosophila |
| 15 | 37.4 | 11.9 | 698 | 12 | BH421278 BH421278 |
| 16 | 37.2 | 11.8 | 417 | 12 | AZ784456 AZ784456 2M0027622 |
| 17 | 37.2 | 11.8 | 588 | 12 | BH538650 BH538650 |

| | | | | | |
|----|------|------|------|----|----------|
| 18 | 36.8 | 11.7 | 563 | 9 | AL587852 |
| 19 | 36.6 | 11.6 | 564 | 12 | AO457327 |
| 20 | 36.6 | 11.6 | 601 | 9 | AL643997 |
| 21 | 36.4 | 11.6 | 394 | 12 | AO66673 |
| 22 | 36.4 | 11.6 | 1086 | 12 | CNS00YXK |
| 23 | 36.2 | 11.5 | 404 | 12 | AO178095 |
| 24 | 36.2 | 11.5 | 421 | 10 | BM162279 |
| 25 | 36.2 | 11.5 | 1041 | 12 | CNS0151D |
| 26 | 36 | 11.4 | 806 | 12 | CNS008OW |
| 27 | 36 | 11.4 | 815 | 12 | BH561753 |
| 28 | 35.8 | 11.4 | 517 | 12 | B69927 |
| 29 | 35.8 | 11.4 | 520 | 12 | BH062502 |
| 30 | 35.8 | 11.4 | 655 | 12 | AG046736 |
| 31 | 35.8 | 11.4 | 690 | 12 | AZ292215 |
| 32 | 35.8 | 11.4 | 710 | 12 | BH502667 |
| 33 | 35.8 | 11.4 | 791 | 9 | AW242927 |
| 34 | 35.8 | 11.4 | 795 | 10 | BG783502 |
| 35 | 35.8 | 11.4 | 960 | 12 | B18793 |
| 36 | 35.8 | 11.4 | 987 | 12 | CNS014PO |
| 37 | 35.8 | 11.4 | 1096 | 12 | CNS05LPJ |
| 38 | 35.6 | 11.3 | 552 | 12 | AO562320 |
| 39 | 35.6 | 11.3 | 875 | 12 | AZ692611 |
| 40 | 35.4 | 11.2 | 544 | 12 | BH042090 |
| 41 | 35.4 | 11.2 | 657 | 9 | BB213292 |
| 42 | 35.2 | 11.2 | 326 | 9 | AJ285573 |
| 43 | 35.2 | 11.2 | 430 | 12 | AO879467 |
| 44 | 35.2 | 11.2 | 604 | 12 | AZ591623 |
| 45 | 35.2 | 11.2 | 979 | 12 | CNS012JO |

ALIGNMENTS

| | | | | | |
|------------|--|-------------|--------------|--------|-----------------|
| RESULT 1 | BF335657 | 563 bp | mrna | linear | EST 22-NOV-2000 |
| LOCUS | RCI-CT0469-160800-011-e04 | CT0469 | Homo sapiens | cdna | mrna sequence. |
| DEFINITION | BF335657 | | | | |
| ACCESSION | BF335657.1 | GI:11306405 | | | |
| VERSION | | | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | 1 (bases 1 to 563) | | | | |
| AUTHORS | Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Balda,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. | | | | |
| TITLE | Shotgun sequencing of the human transcriptome with ORF expressed | | | | |
| JOURNAL | sequence tags | | | | |
| MEDLINE | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) | | | | |
| COMMENT | 20202663 | | | | |
| | Contact: Simpson A.J.G. | | | | |
| | Laboratory of Cancer Genetics | | | | |
| | Ludwig Institute for Cancer Research | | | | |
| | Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil | | | | |
| | Tel: +55-11-2704922 | | | | |
| | Fax: +55-11-2707001 | | | | |
| | Email: asimpson@ludwig.org.br | | | | |
| | This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL | | | | |
| | (http://www.ludwig.org.br/scripts/gethtml2.pl?l=RCI-CT0469-160800-011-e04&t3=2000-08-16&t4=1) | | | | |
| | Seq primer: puc 18 forward | | | | |
| | High quality sequence start: 21 | | | | |
| | High quality sequence stop: 559. | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..563 | | | | |
| | /organism="Homo sapiens" | | | | |

seq primer: 1/
High quality sequence stop: 208.
Location/Qualifiers

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/ULCR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gehtmlml2.pl?c1=RC3&t2=RC3-CT0283>)
201139-011-f12&t3=1999-11-20&t4=1)

Matches 90; Conservative 69; Mismatches 147; Indels 0; Gaps 0;

QY 10 actgaatttaagaataaacttcttggaataagacaataagaaacttgctccacca 69
 Db 605 MMTAAVMTATACAAVMTATATHTVACATMCAVAVATWMTATATATATMHTMT 664
 QY 70 cttgaatatactcttgagatcagcagcctcccttgaggagacatacaagtcagt 129
 Db 665 HYMTATHTATMTATMTAVAHMTAVMYMYACTATATATATATATMTATMTT 724
 QY 130 gtgcctcttggtgagccacccttatatcaagtaggtagacatacaaatltgaaata 189
 Db 725 WYATMTATMCHHTTATATACWCMCAVMTACAAVATCAVACACAVMHTATATTC 784
 QY 190 gattgcacaataaactggagttatggaaacatcagtagaagaaatacaatctcc 249
 Db 785 TMTATVMMMTAVATMTATMTATATATATATATATATATATATATATMT 844
 QY 250 atcccttcaagagatcattacttgcactcaggaataattgtcatgtatatactac 309
 Db 845 ATATMTCMHTATMTATMTATMTATATATATATATATATATATATATAT 904
 QY 310 ttatgc 315
 Db 905 TATATAC 910

RESULT 6
 CDS00DT7/c
 LOCUS
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR8J21 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL075293.1 GI:4948467
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 Location/Qualifiers
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR8J21"
 /note="end : TET3"

BASE COUNT 312 a 157 c 80 g 328 t 224 others
 ORIGIN

Query Match 12.7%; Score 40; DB 12; Length 1101;
 Best Local Similarity 36.6%; Pred. No. 8.6;
 Matches 79; Conservative 36; Mismatches 101; Indels 0; Gaps 0;

QY 2 aaacactgactcagatttaagaataaacttggagaataagacaatacgaatcagtt 61
 Db 1098 AMATAAATMTATMTATMTATMTATMTATMTATMTATMTATMTATMTAT 1039
 QY 62 ctccacactgaatgatatctcttgagatcagcagcctcccttgaggagacatacaag 121
 Db 1038 ATTATATATMTATMTATMTATMTATMTATMTATMTATMTATMTATMTAT 979
 QY 122 tcaattgttgccttcttgagtcaccacttatatcagtagtagtaagtaacatt 181
 Db 978 TTTTWT 919
 QY 182 tgaataagattgcacacataaactgagttat 217
 Db 918 TTTTWTWTATMTATMTATMTATMTATMTATMTATMTATMTATMTATMTAT 883

RESULT 7
 BF295719/c
 LOCUS
 DEFINITION B295719 PB cDNA #17, Tommaso Pace, Marta Ponzi, and Clara Frontali
 Plasmodium berghei cDNA 5', mRNA sequence.

ACCESSION BF295719 GI:13946032
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Plasmodium berghei.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 1 (bases 1 to 652)
 Carlton, J.M., R. and Dame, J.B.
 The Plasmodium vivax and P. berghei gene sequence tag projects
 Parasitol. Today 16 (10), 409 (2000)
 Contact: Dame JB
 Department of Pathobiology, College of Veterinary Medicine
 University of Florida
 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611-0880, USA
 Tel: 352 392 4700
 Fax: 352 392 9704
 Email: damej@mail.vetmed.ufl.edu
 Seq primer: T3.

FEATURES
 source
 Location/Qualifiers
 1..652
 /organism="Plasmodium berghei"
 /strain="ANKA clone HP (gametocyte producer)"
 /db_xref="taxon:3621"
 /clone_lib="PB cDNA #17, Tommaso Pace, Marta Ponzi, and Clara Frontali"
 /dev_stage="asynchronous blood stage"
 /lab_host="Wistar rats"
 /note="Vector: pBluescript II vector DNA, excised from lambda ZAP II.; Site_1: EcoRI; Site_2: XhoI; Total RNA was extracted from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei, grown in Wistar rats to 30% parasitemia and 2-5% gametocytemia. Contaminating host cell contamination estimated to be approximately 5%. PolyA+ RNA was extracted and reverse transcribed using an oligo dt-XhoI primer (lambda ZAP II cDNA cloning kit, Stratagene). Second strand cDNA was made following the manufacturer's protocol. EcoRI adaptors were ligated to the cDNA, and fragments were ligated into EcoRI/XhoI digested vector."

BASE COUNT 289 a 75 c 79 g 206 t 3 others
 ORIGIN

Query Match 12.6%; Score 39.6; DB 10; Length 652;
 Best Local Similarity 48.8%; Pred. No. 11;
 Matches 102; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

[illegible]

REFERENCE 1 (bases 1 to 650)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LIA8555 row: k column: 23
 High quality sequence stop: 564.

FEATURES

source

1. .650
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_1lib="IMAGE:3499150"
 /clone_1lib="NCI_CGAP_Mam6"
 /sex="female, virgin"
 /tissue_type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site_1: Salt;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 193 a 98 c 185 g 174 t

ORIGIN

Query Match 12.3%; Score 38.8; DB 10; Length 650;
 Best Local Similarity 56.2%; Pred. No. 17;
 Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 10 acccaatttaagaataacttttgagaataagaacaatgaatcaagtcttcacca 69
 Db 174 ATTCCACCTTTGGATGATCCAGCATCTTCAGATGAATGCTTTCTTCACATA 115
 QY 70 ctttaagataatccttagagatacagcccttccttagggagacataaagtcagt 129
 Db 114 CTGAACATACCTCAATATACCTTACCCACACACAGCTTGATCAACCATTCAGTGT 55
 QY 130 gtgcctcttg 139
 Db 54 CTGGCCTTG 45

RESULT 13
 AM746159 551 bp mRNA linear EST 19-JUL-2000
 LOCUS AM746159/c
 DEFINITION WS1_39.D05.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION AM746159
 VERSION AM746159.1 GI:7659897
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 551)
 AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt,
 L.H.
 TITLE An EST database from Sorghum: water-stressed plants
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: 77
 High quality sequence start: 42
 High quality sequence stop: 551
 POLY-A=yes.

FEATURES

source

1. .551
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_1lib="water-stressed 1 (WS1)"
 /note="Organ: Mix of 5-week old plants on days 7 & 8 after
 water was withheld; Vector: Lambda Zap; Site_1: XhoI;
 Site_2: EcoRI; The library was made from poly-A RNA in the
 cloning vector lambda Zap II. Clones to be sequenced were
 prepared by mass excision."

BASE COUNT 144 a 107 c 125 g 175 t

ORIGIN

Query Match 12.2%; Score 38.4; DB 9; Length 551;
 Best Local Similarity 55.1%; Pred. No. 21;
 Matches 75; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 176 aaatttgaagaatagattgtcacacataaactgagttatggaacatcagtaaga 235
 Db 528 AACTAGTAAGTAAGTAATTGGCTTACATCAACAGACCTTGACCAATGTACAATCAGG 469
 QY 236 aaatcacacttcaccccttacagagatcattcttgaactcaggaatattgtca 255
 Db 468 AGATCTCACCTCCAGGCTGTAGAGCTCCAGTATTTGCAAGTGACGAAATTGCA 409
 QY 296 tgtgtattatcactt 311
 Db 408 TATTAATATTCACAT 393

RESULT 14
 CNS0161D 1225 bp DNA linear GSS 26-JUL-1999
 LOCUS CNS0161D/c
 DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
 BACN15C18 of Drosophila library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL106171
 VERSION AL106171.1 GI:5620504
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1225)
 Genoscope.
 DIRECT SUBMISSION
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelobAC11.

FEATURES

source

1. .1225
 /organism="Drosophila melanogaster"
 /plasmid="pBelobAC11"

THIS PAGE BLANK (UPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:24:21 : Search time 90.91 Seconds
(without alignments)
2418.238 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895

Sequence: 1 ctaatcgttactgaacagc.....ataataattttaatgac 895

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/lna/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/lna/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/lna/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/lna/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/lna/PTCUS.COMB.seq: *
6: /cgn2_6/ptodata/2/lna/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 18 | 2.0 | 774 | 4 | US-08-998-416-655 |
| 2 | 17 | 1.9 | 1503 | 3 | US-08-889-841B-1 |
| 3 | 17 | 1.9 | 1503 | 3 | US-08-889-841B-4 |
| 4 | 17 | 1.9 | 1545 | 1 | US-08-400-275-17 |
| 5 | 17 | 1.9 | 1990 | 1 | US-08-256-130A-1 |
| 6 | 17 | 1.9 | 2059 | 1 | US-08-256-130A-2 |
| 7 | 17 | 1.9 | 7498 | 2 | US-08-816-693A-1 |
| 8 | 17 | 1.9 | 7498 | 2 | US-08-885-291-1 |
| 9 | 17 | 1.9 | 7498 | 4 | US-09-496-672-1 |
| 10 | 16 | 1.8 | 275 | 1 | US-08-215-084A-2 |
| 11 | 16 | 1.8 | 275 | 1 | US-08-463-212-2 |
| 12 | 16 | 1.8 | 275 | 1 | US-08-463-212-2 |
| 13 | 16 | 1.8 | 318 | 3 | US-09-188-930-15 |
| 14 | 16 | 1.8 | 429 | 1 | US-08-215-084A-6 |
| 15 | 16 | 1.8 | 429 | 1 | US-08-463-212-6 |
| 16 | 16 | 1.8 | 429 | 1 | US-08-463-211-6 |
| 17 | 16 | 1.8 | 900 | 4 | US-09-085-371-5 |
| 18 | 16 | 1.8 | 1333 | 3 | US-08-889-425-3 |
| 19 | 16 | 1.8 | 1403 | 1 | US-07-872-678A-1 |
| 20 | 16 | 1.8 | 1601 | 1 | US-07-968-971A-10 |
| 21 | 16 | 1.8 | 1601 | 1 | US-08-424-406-6 |
| 22 | 16 | 1.8 | 1621 | 1 | US-07-937-609-22 |
| 23 | 16 | 1.8 | 1621 | 4 | US-08-029-170-22 |
| 24 | 16 | 1.8 | 1744 | 1 | US-08-424-406-7 |
| 25 | 16 | 1.8 | 1744 | 1 | US-08-464-523B-7 |
| 26 | 16 | 1.8 | 1810 | 1 | US-08-299-849B-20 |
| 27 | 16 | 1.8 | 1810 | 2 | US-08-142-368A-20 |

| | | | | | | |
|----|----|-----|------|---|-------------------|--------------------|
| 28 | 16 | 1.8 | 1810 | 3 | US-08-967-727-20 | Sequence 20, Appl |
| 29 | 16 | 1.8 | 1810 | 4 | US-08-037-230D-20 | Sequence 20, Appl |
| 30 | 16 | 1.8 | 1850 | 1 | US-08-225-473-3 | Sequence 3, Appl1 |
| 31 | 16 | 1.8 | 2192 | 3 | US-08-942-001-1 | Sequence 1, Appl1 |
| 32 | 16 | 1.8 | 2192 | 4 | US-09-337-386-1 | Sequence 1, Appl1 |
| 33 | 16 | 1.8 | 2238 | 1 | US-07-841-651-1 | Sequence 1, Appl1 |
| 34 | 16 | 1.8 | 2542 | 1 | US-08-441-370-1 | Sequence 1, Appl1 |
| 35 | 16 | 1.8 | 2991 | 3 | US-08-795-430-48 | Sequence 48, Appl1 |
| 36 | 16 | 1.8 | 3060 | 1 | US-08-098-141-1 | Sequence 1, Appl1 |
| 37 | 16 | 1.8 | 3131 | 3 | US-09-035-648-23 | Sequence 23, Appl1 |
| 38 | 16 | 1.8 | 3131 | 4 | US-09-001-951-23 | Sequence 23, Appl1 |
| 39 | 16 | 1.8 | 3839 | 4 | US-09-056-105-14 | Sequence 14, Appl1 |
| 40 | 16 | 1.8 | 4060 | 1 | US-08-308-949A-1 | Sequence 1, Appl1 |
| 41 | 16 | 1.8 | 4584 | 2 | US-08-901-200A-15 | Sequence 15, Appl1 |
| 42 | 16 | 1.8 | 4584 | 2 | US-09-219-391-15 | Sequence 15, Appl1 |
| 43 | 16 | 1.8 | 5183 | 2 | US-08-870-518-7 | Sequence 7, Appl1 |
| 44 | 16 | 1.8 | 6418 | 1 | US-08-480-528A-11 | Sequence 11, Appl1 |
| 45 | 16 | 1.8 | 6418 | 1 | US-08-479-666-11 | Sequence 11, Appl1 |

ALIGNMENTS

RESULT 1
US-08-998-416-655/c
; Sequence 655, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtel, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 655:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 774 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1438UP

US-08-998-416-655

Query Match

2.0%; Score 18; DB 4; Length 774;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;OY 215 cccctccctgtgtgtc 222
|||||
Db 667 CCCCTCCTCTGTCTGTCG 650

RESULT 2

US-08-889-841B-1
; Sequence 1, Application US/08889841B
; GENERAL INFORMATION:
; APPLICANT: Bertman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703C1P
; CURRENT APPLICATION NUMBER: US/08/889,841B
; CURRENT FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: US 60/676,737
; PRIOR FILING DATE: 1996-07-08
; NUMBER OF SEQ. ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ. ID NO 1
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: HIV
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1503)
US-08-889-841B-1

Query Match

1.9%; Score 17; DB 3; Length 1503;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;OY 508 caatttctcaacaatg 524
|||||
Db 765 caatttctcaacaatg 781

RESULT 3

US-08-889-841B-4
; Sequence 4, Application US/08889841B
; GENERAL INFORMATION:
; APPLICANT: Bertman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703C1P
; CURRENT APPLICATION NUMBER: US/08/889,841B
; CURRENT FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: US 60/676,737
; PRIOR FILING DATE: 1996-07-08
; NUMBER OF SEQ. ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ. ID NO 4
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: HIV
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1503)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1503)
; OTHER INFORMATION: n = A,T,C or G
US-08-889-841B-4Query Match
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;OY 508 caatttctcaacaatg 524
|||||
Db 765 caatttctcaacaatg 781

RESULT 4

US-08-400-275-17
; Sequence 17, Application US/08400275
; Patent No. 5668295
; GENERAL INFORMATION:
; APPLICANT: Wahnab, Samir Z.
; APPLICANT: Malik, Vedpal S.
; TITLE OF INVENTION: PUTRESCINE N-METHYLTRANSFERASE,
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING PUTRESCINE
; TITLE OF INVENTION: N-METHYLTRANSFERASE, AND TRANSGENIC TOBACCO PLANTS WITH
; TITLE OF INVENTION: ALTERED NICOTINE CONTENT
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Ave. of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,275
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,681
; FILING DATE:
; APPLICATION NUMBER: US 07/613,160
; FILING DATE: 14-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 34,310
; REFERENCE/DOCKET NUMBER: PM-1696
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ. ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1545 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-400-275-17Query Match
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;OY 844 aaatgcaatcaggaag 860
|||||
Db 1227 AATGCAATCAGGAAG 1243

RESULT 5

US-08-256-130A-1
; Sequence 1, Application US/08256130A
; Patent No. 5659121
; GENERAL INFORMATION:
; APPLICANT: BIRD, COLIN R.Query Match
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

; APPLICANT: GRIERSON, DONALD
; APPLICANT: HALL, LISA N.
; TITLE OF INVENTION: DNA, DNA CONSTRUCTS, CELLS AND PLANTS
; TITLE OF INVENTION: DERIVED THEREFROM
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/256.130A
; FILING DATE: 30-SEP-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9127098.3
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT.GB92.02355
; FILING DATE: 18-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 205427/SEE 36697/UST
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1990 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-256-130A-1

Query Match 1.9%; Score 17; DB 1; Length 1990;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 710 ttgtttgtgtctct 726
|||||
Db 102 TTTGTTGTGCTCT 118

RESULT 6
US-08-256-130A-2
; Sequence 2, Application US/08256130A
; Patent No. 5659121
; GENERAL INFORMATION:
; APPLICANT: BIRD, COLIN R.
; APPLICANT: GRIERSON, DONALD
; APPLICANT: HALL, LISA N.
; TITLE OF INVENTION: DNA, DNA CONSTRUCTS, CELLS AND PLANTS
; TITLE OF INVENTION: DERIVED THEREFROM
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/256.130A
; FILING DATE: 30-SEP-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9127098.3
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT.GB92.02355
; FILING DATE: 18-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 205427/SEE 36697/UST
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2059 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-256-130A-2

```

```

Query Match 1.9%; Score 17; DB 1; Length 2059;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 710 ttgtttgtgtctct 726
|||||
Db 99 TTTGTTGTGCTCT 115

```

```

RESULT 7
US-08-816-693A-1/c
; Sequence 1, Application US/08816693A
; Patent No. 5874241
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S
; APPLICANT: Turek, Fred W
; APPLICANT: Pinto, Lawrence H
; TITLE OF INVENTION: Clock Gene and Gene Product
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Rocky, Milnamow & Katz
; STREET: Two Prudential Plaza, Suite 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816.693A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5874241thrup, Thomas E
; REGISTRATION NUMBER: 33,268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5400
; TELEFAX: 312-616-5460

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7498 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 389..2954
; US-08-816-693A-1

Query Match 1.9%; Score 17; DB 2; Length 7498;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ctcaaatgggaacgctg 66
|||||
Db 7255 CTCGAATGGGAACGCTG 7239

RESULT 8
US-08-885-291-1/c
; Sequence 1, Application US/08885291A
; Patent No. 6057125
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/08/885,291A
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 08/816,693
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 7498
; TYPE: DNA
; ORGANISM: Mus musculus
; US-08-885-291-1

Query Match 1.9%; Score 17; DB 3; Length 7498;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ctcaaatgggaacgctg 66
|||||
Db 7255 CTCGAATGGGAACGCTG 7239

RESULT 9
US-09-496-672-1/c
; Sequence 1, Application US/09496672
; Patent No. 6291429
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/09/496,672
; CURRENT FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 08/885,291
; PRIOR FILING DATE: 1997-06-30
; PRIOR APPLICATION NUMBER: 08/816,693
; PRIOR FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1

; LENGTH: 7498
; TYPE: DNA
; ORGANISM: Mus musculus
; US-09-496-672-1

Query Match 1.9%; Score 17; DB 4; Length 7498;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ctcaaatgggaacgctg 66
|||||
Db 7255 CTCGAATGGGAACGCTG 7239

RESULT 10
US-08-215-084A-2/c
; Sequence 2, Application US/08215084A
; Patent No. 5461032
; GENERAL INFORMATION:
; APPLICANT: Karen J. Krapcho, Bradford Carr Van
; APPLICANT: Magenen, J.R. Hunter Jackson and Robert
; APPLICANT: M. Kral, Jr.
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE
; TITLE OF INVENTION: PEPTIDES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESSEE: and No. 5461032rls
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/215,084A
; FILING DATE: March 18, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 859,925
; FILING DATE: March 24, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Caldwell, Esq.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: FMC-0069
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-215-084A-2

Query Match 1.8%; Score 16; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 aatccagggtgcgca 283
|||||
Db 115 AATCCAGGGGTGCCA 100

RESULT 11
US-08-463-212-2/c

```
; Sequence 2, Application US/08463212
; Patent No. 5658563
; GENERAL INFORMATION:
; APPLICANT: Karen J. Krapcho, John Randolph Hunter
; APPLICANT: Jackson, Bradford C. VanWagenen, Robert
; APPLICANT: M. Kral, Jr.
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE
; TITLE OF INVENTION: PEPTIDES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESSEE: and No. 5658563rls
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 diskette, 720 Kb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,212
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/215,084
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Caldwell, Esq.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: FWC-0076
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-463-212-2
;
; Query Match 1.8%; Score 16; DB 1; Length 275;
; Best Local Similarity 100.0%; Pred. No. 92;
; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 268 aatcaggggtcgcca 283
; DB 115 AATCAGGGGTCCGCA 100
;
; RESULT 12
; US-08-463-211-2/c
; Sequence 2, Application US/08463211
; Patent No. 5658781
; GENERAL INFORMATION:
; APPLICANT: Karen J. Krapcho, John Randolph Hunter
; APPLICANT: Jackson, Bradford C. VanWagenen, Robert
; APPLICANT: M. Kral, Jr.
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE
; TITLE OF INVENTION: PEPTIDES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESSEE: and No. 5658781rls
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 diskette, 720 Kb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,211
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/215,084
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Caldwell, Esq.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: FWC-0075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-463-211-2
```

```
Query Match 1.8%; Score 16; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 268 aatcaggggtcgcca 283
DB 115 AATCAGGGGTCCGCA 100
```

```
; RESULT 13
; US-09-188-930-15
; Sequence 15, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Rat
;
; US-09-188-930-15
```

```
Query Match 1.8%; Score 16; DB 3; Length 318;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 351 caaagatgttctcgt 366
DB 24 caaagatgttctcgt 39
```

```
; RESULT 14
; US-08-215-084A-6/c
; Sequence 6, Application US/08215084A
```

```
; Patent No. 5461032
; GENERAL INFORMATION:
; APPLICANT: Karen J. Krapcho, Bradford Carr Van
; APPLICANT: Wagenen, J.R. Hunter Jackson and Robert
; APPLICANT: M. Kral, Jr.
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE
; TITLE OF INVENTION: PEPTIDES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESS: and No. 5461032r1s
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/215,084A
; FILING DATE: March 18, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 859,925
; FILING DATE: March 24, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Caldwell, Esq.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: FMC-0069
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-215-084A-6
```

```
Query Match 1.8%; Score 16; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 268 aatccagggtcgcca 283
      ||||||||||||
Db 288 AATCCAGGCGTCGCCA 273
```

```
RESULT 15
US-08-463-212-6/c
; Sequence 6, Application US/08463212
; Patent No. 5658563
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Karen J. Krapcho, John Randolph Hunter
; APPLICANT: Jackson, Bradford C. VanWagenen, Robert
; APPLICANT: M. Kral, Jr.
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE
; TITLE OF INVENTION: PEPTIDES
; NUMBER OF SEQUENCES: 8
```

```
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESSEE: and No. 5658563r1s
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
```

```
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: 3.5 diskette, 720 Kb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,212
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/215,084
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Caldwell, Esq.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: FMC-0076
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-463-212-6
```

```
Query Match 1.8%; Score 16; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 268 aatccagggtcgcca 283
      ||||||||||||
Db 288 AATCCAGGCGTCGCCA 273
```

```
Search completed: September 16, 2002, 06:24:30
Job time: 10643 sec
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 05:18:16 : Search time 3017.73 Seconds
(without alignments)
4002.934 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895

Sequence: 1 ctactcgttactgtaacgc.....aataaatttttaaatgctc 895

Scoring table:

OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estbta:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 521 | 58.2 | 613 | 10 | BG536626 602566231 |
| 2 | 497 | 55.5 | 498 | 10 | BF196945 7180f03.x |
| 3 | 455 | 50.8 | 491 | 9 | AT340056 qn18h09.x |
| 4 | 426 | 47.6 | 426 | 9 | AA479746 zu35g03.s |
| 5 | 398 | 44.5 | 504 | 9 | AI809681 wf76g06.x |
| 6 | 397 | 44.4 | 397 | 9 | AA481852 zva2f04.r |
| 7 | 355 | 39.7 | 477 | 9 | AT142875 ox64e06.s |
| 8 | 355 | 39.7 | 477 | 9 | BE148882 CM2-HT024 |
| 9 | 339 | 37.9 | 431 | 9 | AT079465 oz39f05.s |
| 10 | 334 | 37.3 | 820 | 10 | BI770944 603059769 |
| 11 | 316 | 35.3 | 419 | 9 | AA430487 zw23f06.s |
| 12 | 315 | 35.2 | 429 | 9 | AA298444 UI-H-BMO- |
| 13 | 297 | 33.2 | 491 | 9 | AA317141 ESI19048 |
| 14 | 269 | 30.1 | 491 | 9 | AT830551 wj51b02.x |
| 15 | 267 | 29.8 | 270 | 9 | AT860310 w101e10.x |
| 16 | 267 | 29.8 | 461 | 9 | AT276107 q174h09.x |
| 17 | 266 | 29.7 | 356 | 9 | AW770737 h189g02.x |

| | | | | | | |
|------|-----|------|------|----|----------|--------------------|
| C 18 | 263 | 29.4 | 263 | 10 | BE645174 | BE645174 7e64d07.x |
| C 19 | 263 | 29.4 | 263 | 10 | BF055365 | BF055365 7j78f09.x |
| C 20 | 263 | 29.4 | 408 | 9 | AM183010 | AM183010 xp99f12.x |
| C 21 | 262 | 29.3 | 667 | 9 | AW471176 | AW471176 xv13a03.x |
| C 22 | 261 | 29.2 | 482 | 10 | BF061539 | BF061539 7j52g10.x |
| C 23 | 257 | 28.7 | 326 | 10 | BF882906 | BF882906 QV3-ET020 |
| C 24 | 256 | 28.6 | 404 | 10 | BF745581 | BF745581 QV0-BT084 |
| C 25 | 254 | 28.4 | 255 | 9 | AT207071 | AT207071 qf80f02.x |
| C 26 | 250 | 27.9 | 411 | 9 | AT368232 | AT368232 qw16e02.x |
| C 27 | 238 | 26.6 | 342 | 9 | AA477092 | AA477092 zu35g03.x |
| C 28 | 238 | 26.6 | 392 | 9 | AA433968 | AA433968 zw23f07.x |
| C 29 | 234 | 26.1 | 251 | 10 | BF743524 | BF743524 QV0-BT084 |
| C 30 | 208 | 23.2 | 254 | 9 | AM389442 | AM389442 IL0-ST016 |
| C 31 | 207 | 23.1 | 409 | 9 | BE073132 | BE073132 MR3-BT055 |
| C 32 | 205 | 22.9 | 410 | 10 | BF745586 | BF745586 QV0-BT084 |
| C 33 | 191 | 21.3 | 271 | 9 | BE181362 | BE181362 QV0-HT063 |
| C 34 | 188 | 21.0 | 232 | 9 | AT474243 | AT474243 tg84c11.x |
| C 35 | 181 | 20.2 | 422 | 10 | BF743581 | BF743581 QV0-BT084 |
| C 36 | 177 | 19.8 | 302 | 9 | AM770575 | AM770575 h186g07.x |
| C 37 | 159 | 17.8 | 223 | 9 | AM797203 | AM797203 QV1-DM003 |
| C 38 | 155 | 17.3 | 356 | 9 | AM173795 | AM173795 xp98f10.x |
| C 39 | 135 | 15.1 | 144 | 9 | AA074399 | AA074399 zm16f08.s |
| C 40 | 133 | 14.9 | 267 | 10 | BE743520 | BE743520 QV0-BT084 |
| C 41 | 130 | 14.5 | 279 | 9 | AT758852 | AT758852 ty16c01.x |
| C 42 | 122 | 13.6 | 401 | 9 | AT276273 | AT276273 q165g05.x |
| C 43 | 117 | 13.1 | 192 | 10 | CO0521 | CO0521 HMG5000804 |
| C 44 | 115 | 12.8 | 1279 | 10 | BG535485 | BG535485 602563170 |
| C 45 | 114 | 12.7 | 227 | 9 | AT680037 | AT680037 tu24c01.x |

ALIGNMENTS

RESULT 1
BG536626 613 bp mRNA linear EST 03-APR-2001
LOCUS 602566231F1 NIH_MGC_77 Homo sapiens CDNA clone IMAGE:4691118 5',
DEFINITION mRNA sequence.
ACCESSION BG536626
VERSION BG536626.1 GI:13528159
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M1509 row: m column: 07
High quality sequence stop: 611.
Location/Qualifiers
1. 613
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4691118"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1:
SfiI (ggcgcctcgcc); Site: 2: SfiI (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCATGAGCCGAGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9

kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 165 a 176 c 132 g 140 t

ORIGIN

Query Match 58.2%; Score 521; DB 10; Length 613;
Best Local Similarity 99.8%; Pred. No. 1.3e-260;
Matches 571; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 140 gcagggaaccttcattatatttccaagaacttaccgtgcagcaggttcagatga 199
DB 14 GCAGGGACCTTCATTAATATTTTCAAGCAACTTACAGCTGCACCCAGATTCCGATGA 73
QY 200 aatttcaactcttcctctctctctgtgtgtgcactaatgtatgttcattctcta 259
DB 74 AAGTTTAACTCTCTCCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 133
QY 260 gcagcctgaaatccagggttcgcaggaagcagcagcagcagcagcagcagcagcagc 319
DB 134 GCAGCCTGAATTCAGAGGGGTGCGCAGAGGCCAGAGGCCAGGCTTTAGAGAGAT 193
QY 320 ggtctccaggaagcgcgcgaagaatgtgagtgcaaaagattgttcttcagagcccgagaa 379
DB 194 GGTCTCCAGAGAGGGCGGCAAGAAATGTGATGCAAAAGATTGCTTCCTGAGAGCCCGAGAA 253
QY 380 gaaatcatcagacagtgctcgtgcgcgaagaagcagtgccctgtgcatcattcaag 439
DB 254 GAAATTCATGACAGATGCTGCGGCTGCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 313
QY 440 gcaatgtgaaagaaac 499
DB 314 GCAATGTGAAGAAACACACACACCAAGGACACACAGAAAGCCAAACACATTTCCAGAG 373
QY 500 cctgcgaagcaatttccaacaatgtcagtaagaagccttgccttcgttcagagc 559
DB 374 CCGCCAGCAGCAATTTCTCAACCAATGTGAGTAAGAACTTGGCTTCTGTAAGAGC 433
QY 560 tctgagcgcacactcttccaatttaaacattctcagcacaagaagacagtgagcacaactac 619
DB 434 TCTGAGGGCCCTCTTCCCAATTAACATTTCTGACGCCAAGACAGATGAGCACCTTAC 493
QY 620 cagacactcttcttccaactcactcactcactcactcactcactcactcactcactcactc 679
DB 494 CAGACACTCTTCTTCCCTCCACTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 553
QY 680 gctctcaaaaagatgttttcaaatcaatt 711
DB 554 GCTCTCAAAAAGCATGTTTTCAGATCAATTT 585

RESULT 2
BF196945/c 498 bp mRNA linear EST 03-NOV-2000
LOCUS 7180F03.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:3529/932 3', mRNA sequence.
ACCESSION BF196945
VERSION BF196945.1 GI:11085527
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 498)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.jnl.gov) for further information.
High quality sequence stop: 487.
Location/Qualifiers

FEATURES
Source

1.498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3527932"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pUT73P-Pac (Pharmacia) with a modified polylinker; Site:1: Not 1; Site:2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHP pool 1: 309384-310919, 323208-325895 Soares NBHP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NBHPF-9W pool 1: 758280-760583, 772104-774407 Soares NBHP pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHP pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 132 a 83 c 138 g 145 t

ORIGIN

Query Match 55.5%; Score 497; DB 10; Length 498;
Best Local Similarity 100.0%; Pred. No. 4.3e-248;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 tgggctgcgaagaagcagcagcctgtgcatcttccaagggaatgtgaagaagaacag 458
DB 498 TGGGCTGCCAAGAAAGCAGTGCCTCTGTGATCTTTCAAGGCAATGTGAAGAAACAG 439
QY 459 acaccaagcgcacacagaagaagcacaacacacacacacacacacacacacacacacacac 518
DB 438 ACACCAAGGACACACAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 379
QY 519 acaatgtcagtaagaagcttgccttcgttcgttcgttcgttcgttcgttcgttcgttcgttc 578
DB 378 ACAATGTCAAGTAAGAAAGCTTGTCTGCTTGTGTGAGAGCTCTGAGCGCCACTCTTCC 319
QY 579 aattaaactctcagaagaagacagtgagcaccctcagaactcttctcc 638
DB 318 AATTAACATTTCTCAGCAGAAAGACAGTACACACTTCTTCTTCTCC 259
QY 639 acctcaactctccactgtlaccacccctaataatcactcagtgctcctcaaaaagcatgtt 698
DB 258 ACCTCACTCTCCACAGTACCCACCCCTTAATCATTCAGTGTCTCAAAAACCATGTTT 199
QY 699 ttcaaatatatttgtttgttctctctcagtgcttcttctcgttcagtccttagct 758
DB 198 TTCAAGATCATTTTGTGCTCTCTCTAGTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 139
QY 759 gtgcccctccctaccagcagcgttaacttaacttgaagaattccaggaactttagc 818
DB 138 GTGCCCTCCCTTACCAGAGGCTTAGGCTTAATTAAGCAAGTTCAGGAAGAACTGTAC 79
QY 819 ttccctagctagtgtaatttaaccttaaatgcaatcagaagaatgacagaagtcaat 878
DB 78 TTCCCTAGCTAGTGTCTTAACTTAACCTTAATGCAATCGAAGAGCAAGCAAGCAAGCAAT 19
QY 879 aatatattttaaattgc 895
DB 18 AATATATTTTAAATGTC 2

RESULT 3
A1340056/c 491 bp mRNA linear EST 13-FEB-1999
LOCUS A1340056

| DEFINITION | 9018h09.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1898657 3', mRNA sequence. |
|---------------------------|--|
| ACCESSION | A1340056 |
| VERSION | A1340056.1 |
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. |
| AUTHORS | 1 (bases 1 to 491) |
| TITLE | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . |
| JOURNAL | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index |
| COMMENT | Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps@email.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/bdnp/image/image.html Insert Length: 485 Std Error: 0.00 Seq primer: -400P from Glbco High quality sequence stop: 444. Location/Qualifiers 1..491 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1898657" /clone_1db="NCI_CGAP_Lu5" /tissue_type="carcinoid" /lab_host="DH10B" /note="Organ: lung; Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo." |
| BASE COUNT | 101 a 107 c 141 g 142 t |
| ORIGIN | |
| Query Match | 50.8%; Score 455; DB 9; Length 491; |
| Best Local Similarity | 100.0%; Pred. No. 3,6e-226; |
| Matches 455; Conservative | 0; Mismatches 0; Indels 0; Gaps 0; |
| 219 | cctcctgttggtgcacatgctcgaigtgcatagtctctagcagcctgtaatccagggt 278 |
| 491 | ctcctctgttggtgcacatgctcgaigtgcatagtctctagcagcctgtaatccagggt 432 |
| 279 | cgcacagagccacagagaccgagccagaggtctctagagagatgctcgaagagagcgccca 338 |
| 431 | cgccagagagccacagagaccgagccagaggtctctagagagatgctcgaagagagcgccca 372 |
| 339 | agaatgtagatgcaaaagtgtgtctctgagagagcccgagaaataatcatgacagtgtc 398 |
| 371 | agaaatgtagatgcaaaagtgtgtctctgagagagcccgagaaataatcatgacagtgtc 312 |
| 399 | tgagctgccaaagagcagtgccctctgacatcttcaagaagcaatgtgaagaaacaag 458 |
| 311 | tggcgctgcacaaagacagtgctccctctgtgattcatttcacaggccatgtgaagaaacag 252 |
| 459 | acaccaaagagccacacagaaagccaacaagcatctccagagcctgcgcagcaattctcaa 518 |
| 251 | acaccaaagagccacacagaaagccaacaagcatctccagagcctgcgcagcaattctcaa 192 |
| 519 | acaaatgtcagtaagaagcttgcctctgccttctgtagagctctgagcgcctcttc 578 |

| | | | |
|-----------------------|---|--|-----------------------------|
| Db | 191 | ACAATGTCAGCTAAGAGACTTGTCTGCTTGTAGAGACTGTGAGGCCACACTCTTCC | 132 |
| Qy | 579 | aattaaactcttcgagcaagaagacagtgagacacaccacagacactctctcc | 638 |
| Db | 131 | AATTAAACTTCTCAGCCCAAGAAAGACAGTGACACACCTTACACGACACTCTTCTCC | 72 |
| Qy | 639 | acctcaactctccactgttaccaccccttaacat | 673 |
| Db | 71 | ACCTCACTCTCCACTGTACACCCACTTAATCAT | 37 |
| RESULT | 4 | | |
| LOCUS | AA479746/c | 426 bp | mRNA linear EST 08-AUG-1997 |
| DEFINITION | zuz5g03.s1 Soares ovary tumor NbhOT Homo sapiens cDNA clone | | |
| ACCESSION | IMAGE:740020 3', mRNA sequence. | | |
| VERSION | AA479746 | | |
| KEYWORDS | AA479746.1 GI:2205632 | | |
| SOURCE | EST. | | |
| ORGANISM | human. | | |
| REFERENCE | Homo sapiens | | |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 426) Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacey, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R. Washu-Merck EST Project 1997 Unpublished (1997) | | |
| TITLE | Unpublished (1997) | | |
| JOURNAL | Contact: Wilson RK | | |
| COMMENT | Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seg primer: -41m3 fwd. ET from Amerham High quality sequence stop: 419. | | |
| FEATURES | Location/Qualifiers | | |
| SOURCE | 1. 426 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="GDB:5939727" | | |
| | /db_xref="taxon:9606" | | |
| | /clone="IMAGE:740020" | | |
| | /clone_lib="Soares ovary tumor NbhOT" | | |
| | /sex="Female" | | |
| | /tissue_type="Ovarian tumor" | | |
| | /lab_host="DH10B (ampicillin resistant)" | | |
| | /note="Organ: ovary; Vector: pRT3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGAGGCGGCCGCCGGTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT3 vector (Pharmacia). Library constructed by Bento Soares and M.Felima Bonaldo." | | |
| BASE COUNT | 121 a 64 c 122 g 119 t | | |
| ORIGIN | | | |
| Query Match | 47.6%; | Score 426; | DB 9; |
| Best local Similarity | 100.0%; | Pred. NO. 4.7e-211; | |
| Matches 426; | Conservative 0; | Mismatches 0; | Indels 0; |
| Gaps 0; | | | |
| Db | 469 | caccacagaaacccaagaacatccagagcgtcgacgaacttcctcaacatgctcag | 528 |
| Db | 426 | CACGACAGAAAGCCAAACAGCATCTCCAGAGCTCCACACATTTCTCAACAAATGTCA | 367 |
| Qy | 529 | ctaagaagccttgcctcgtccttcttgtagagcctcgtagcgccacacttcattaaacat | 588 |

Db 366 CTGAAGACTTTGCTGCTGCTTTGTAGAGCTCTGAGCGCCCACTCTTCCATTAAACAT 307
 QY 589 tctcagcaagaagacagtagcaccacacacacacacactctctctccacacacactctc 648
 Db 306 TCTCAGCGCAAGAGAGAGTAGGACACACTTACGACACTCTTCTCTCCACCTCACTCT 247
 QY 649 cccactgtacccaccccttaataatcctcagtgctctcaaaaagcagcttttccaagta 708
 Db 246 CCCACTGTACCCACCCCTTAATCATTCAGTGTCTCAAAAAGCATGTTTTCAAGATCA 187
 QY 709 ttgt 768
 Db 186 TTTTGT 127
 QY 769 ctacccagccttagccttaataatcctcgaagaatcctcgaagaactgtagcttccagta 828
 Db 126 CTTACCCAGGCTTAGGCTTATTACCTGAAGAAGATCCAGGAACCTGTAGCTTCTACTA 67
 QY 829 gtgtcatttaaccttaaatgcaatcagaagaagtagcaagaagatcaataatatttt 888
 Db 66 GTGTCAATTTAACTTAATGCAATCAGGAAGTAGCAAAAGCAAGTCAATTAATATTTT 7
 QY 889 aatgt 894
 Db 6 AAATGT 1
 RESULT 5 504 bp mRNA linear EST 07-JUL-1999
 AI809681
 LOCUS wf76c06.x1 Soares_NFL_T_GBC_SI Homo sapiens cDNA clone
 DEFINITION IMAGE:2361514 3', mRNA sequence.
 ACCESSION AI809681
 VERSION AI809681.1 GI:5396247
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 504)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 459.
 FEATURES
 source
 1..504
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2361514"
 /clone_1lb="Soares_NFL_T_GBC_SI"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site: 1; Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBH19W, testis NHT, and B-cell
 NCI-CGAP-GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 1.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo. "
 BASE COUNT 134 a 85 c 141 g 144 t
 ORIGIN

Query Match 44.5%; Score 398; DB 9; Length 504;
 Best Local Similarity 99.8%; Pred. No. 2e-196;
 Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 407 caaagaagcagtgccctgtgcatcattcaagggcaatgtggaagaacaagaccacaa 466
 Db 490 CAAAGAAGCAGTGCCCTGTGATCATTTCAAGGCAATGTGAAGAAACAAAGACACCAA 431
 QY 467 ggcaccacagaagaagcccaaacatccagagcctcgcagcaaatctcaacaatgtc 526
 Db 430 GGCACCAAGAAAGCCAGAGACATTCAGAGCTGCAAGCAATTTCTAACAATGTC 371
 QY 527 agctaaagaagcttgctcgtccttltgaagagctctgaagccacactctccaattaac 586
 Db 370 AGCTAAGAAGCTTTGGCTGCTGCTTTGAGAGAGCTGAGAGGCCACCTTCAATTAAAC 311
 QY 587 attctagccaagaagaagcagtagcagacactaccagaacactctctctccacactact 646
 Db 310 ATTCTCAGCCAGAGAGACAGTGAACACCTTACAGACACTCTTCTTCCACCTCACT 251
 QY 647 ctccacatgtaccacccccaataatcattccagtgctctcaaaaagcagatgtttcaaat 706
 Db 250 CTCCCACTGTACCCACCCCTAAATCATTCAGAGCTCTCAAAAAGCATGTTTCAAGAT 191
 QY 707 catttggt 766
 Db 190 CATTTGTTGTTGCTGCTCTCTAGTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 131
 QY 767 cccctaccagcgttagccttaataatcactgaagaagatccagaagaactgtagctctagc 826
 Db 130 CCTTACCCAGGCTTAGGCTTAATTAATCCTGAAGATTCAGGAACCTGTAGCTTCTTAGC 71
 QY 827 tagtgcatttaaccttaaatgcaatcag 855
 Db 70 TAGTGCATTTAACTTAATGCAATCAG 42
 RESULT 6
 AA481852 397 bp mRNA linear EST 08-AUG-1997
 LOCUS zva2f04.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
 DEFINITION IMAGE:756319 5', mRNA sequence.
 ACCESSION AA481852
 VERSION AA481852.1 GI:2209530
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 397)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S.,
 Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
 Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie,
 T., Waterston,R. and Wilson,R.
 WASHU-MERCK EST Project 1997
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilison RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -28m13 rev2 ET from Amersham.
 FEATURES
 source
 1..397
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:756319"
 /clone_1lb="Soares ovary tumor NbHOT"
 /sex="Female"
 /tissue_type="Ovarian tumor"

```

BASE COUNT      107 a      107 t
ORIGIN
107 a      121 c      62 g      107 t
M.Fatima Bonaldo.
(Pharmacia). library constructed by Bento Soares and
the Not I and Eco RI sites of a modified p773 vector
adapters (Pharmacia), digested with Not I and Eco RI
double-stranded cDNA was size selected, ligated to Eco RI
primers (5' TGTTACCAATCTGTAAGTGGAGCGCGCCGGTTTTTTTTTTTTTTT 3'),
strand cDNA was primed with a Not I - oligo(dT) primer (5'
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
/ab host="DH10B (ampicillin resistant)"

```

| | | | | |
|-----------------------|----------------|--------------|----------|------------|
| Query Match | 44.4% | Score 397 | DB 9 | Length 397 |
| Best Local Similarity | 100.0% | Pred. No. | 6.4e-196 | |
| Matches 397 | Conservative 0 | Mismatches 0 | Indels 0 | Gaps 0 |

[illegible]

QY 505 cagcaattctccaacacatgctcagctaagaagccttgctctgcgccttgytaagagctctga 56
|||||
Db 61 CAGCAATTTTCCAACAATGTGACGTAAAGAAGCTTTGGCTCTGCGCTTTGTAGGACCTCTGA 12

Oy 565 gcgcccactcttcacatatacaattcgcagcaagaagacagttagcacacctaaccagac 62
|||||
|||
Db 121 GCGCCACCTTTCACATTAACAATTCTCAGGCCAAGAAGACAGTGAGCACACCTACCAGAC 180
|||||

07 625 acctctctctccacactcactctcccaatgtaccacccctaataatccagtgctc 68
|||||
Db 181 accttcttctccacactcactctcccaatgtaccacccctaataatccagtgctc 24
|||||

| | |
|----|--|
| Oy | 685 caaaagcagtgttttcaagatcatctggttgctctcttagtcgtctctctc 74 |
| Db | 241 CAAAAGCAGCTTTTCAAGATCATTTGTTGTCTCTCTAAGTCCTTCTCTC 30 |

QY 745 gtcagctcttagcctgtgcctccctaccagcttagcttaattacctaagaattc 80
|||||
Db 301 GTAGACTTTCAGCCTGTGCCCCCTCCCTTACCCAGCCTTAGCCTTAATTACCTGAAGATTCC 36

Qy 805 caggaacigttagcttcctagctagtgatcattaac 841
|||||
Db 361 CAGGAACCTGTAGCTTCCTAGCTAGTGATCATTAACT 397

RESULT 7
AT142875/c

| | | | | | |
|------------|---|--------|-------|--------|-----------|
| LOCUS | AI142875 | 477 bp | miRNA | linear | ESI 23-06 |
| DEFINITION | ox6406.s1 Soares_NhmPu_s1 Homo sapiens cdna clone IMAGE:16637, mRNA sequence. | | | | |
| ACCESSION | AI142875 | | | | |

| | |
|----------|--------------|
| ORGANISM | Homo sapiens |
| SOURCE | human. |
| KEYWORDS | EST. |
| VERSION | AL142872.1 |
| GENEID | GI:5039254 |

REFERENCE
AUTHORS
MOL-CCGAP <http://www.ncbi.nlm.nih.gov/molccgap>
1 (bases 1 to 477)
Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.

| | |
|---------|---|
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP) |
| JOURNAL | Tumor Gene Index |
| COMMENT | Unpublished (1997) |
| CONTACT | Robert Strussberg, Ph.D. |

Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information
Recent research 800 543-7777 0 00

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 383.
Location/Qualifiers
1 477

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:166112"
/clone_1db="Source_NbHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pRTT3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHMP, pregnant uterus
NbHMPu, and fetal heart NbH119W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,

```

| BASE COUNT ORIGIN | 129 a | 74 c | 135 g | 139 t |
|----------------------|-------|------|-------|-------|
|----------------------|-------|------|-------|-------|

| | | | | |
|---------------------------|--------|--------------------|-------|-------------------|
| Query Match | 39.7% | Score 355; | DB 9; | Length 477; |
| Best Local Similarity | 99.8%; | Pred. No.5.5e-174; | | |
| Matches 475; Conservative | 0; | Mismatches | 0; | Indels 1; Gaps 1; |

| | | |
|-----|--|-----|
| QY | ccccctgatacatlccaagggcaatgtgaagaataaacacagaccaccaaaaggcaccacagaaa | 479 |
| Db | cccttgatcatatttcacagggcaatgtgaagaataaacacagaccaccaaaaggcaccacagaaa | 478 |
| 477 | ccccctgatacatlccaagggcaatgtgaagaataaacacagaccaccaaaaggcaccacagaaa | 479 |

| | | | |
|----|-----|--|-----|
| QY | 480 | gccaaacaagcattccagagcctgcccagcaatttctcaaacatgctagcctagaagctt | 539 |
| Db | 417 | gccaaacagcatttccagagcctgccagcaattttctaaacatgctagcctagaagctt | 358 |

| QY | 540 | 599 |
|----|---|-----|
| QY | tgccgagccttgtagagccttgagcgccacctcttcacatlaaacatcttcagccaag | 599 |
| Db | 357 T -CTCTCCCTTGTAGAGCTCTGAGCGCCCTCTTCCCAATTAACATTCTCAGCCAAG | 299 |

| | | | | | |
|----|-----|------------------------------|------------------------------|------|-----|
| QY | 600 | aagacagtgagcacaccctaccagacac | ctctctctccacactcaactctccacgt | aac | 659 |
| Db | 298 | AAGACAGTGAGCACACCCTACCAGACAC | CTCTCTCTCCACACTCACTCTCCACTG | TACC | 239 |

660 caccctaatcatctccagtgctctcctcaaaagcatglttlltccaagatcaatltgttctgtc 719

Dp 238 cacccttaaatctatttccagtcgtctctcaaaagcatgtttttcgaagatcatgtttgtgttt 179

Dp 720 gctcctatagtgcttcttcctcgcgtaagtcttagccgtgtagccctcccctaaccaaggc 779
|||||
178 gcctctctctagatgctctttctctcctgcacgtctttagcctgtgccctcccccttacccaggg 119
|||||

Oy 780 tttagcctaattacctgaagaattccagaaactgtagcttccttagctagtgatcattaa 839
 Db 118 tttagggcttaattaaccttaagaattccaggaactgtgagcttctcttagctgacattttta 59

| | | |
|-----|--|-----|
| 840 | cctlaaactgcaatcaggaagttagcaacagagtccaataatattttaaattgc | 895 |
| 841 | | |
| 842 | | |
| 843 | | |
| 844 | | |
| 845 | | |
| 846 | | |
| 847 | | |
| 848 | | |
| 849 | | |
| 850 | | |
| 851 | | |
| 852 | | |
| 853 | | |
| 854 | | |
| 855 | | |
| 856 | | |
| 857 | | |
| 858 | | |
| 859 | | |
| 860 | | |
| 861 | | |
| 862 | | |
| 863 | | |
| 864 | | |
| 865 | | |
| 866 | | |
| 867 | | |
| 868 | | |
| 869 | | |
| 870 | | |
| 871 | | |
| 872 | | |
| 873 | | |
| 874 | | |
| 875 | | |
| 876 | | |
| 877 | | |
| 878 | | |
| 879 | | |
| 880 | | |
| 881 | | |
| 882 | | |
| 883 | | |
| 884 | | |
| 885 | | |
| 886 | | |
| 887 | | |
| 888 | | |
| 889 | | |
| 890 | | |
| 891 | | |
| 892 | | |
| 893 | | |
| 894 | | |
| 895 | | |

RESULT 8
0014000

| | | | | | |
|------------|---------------------------|--------|--------------|--------|-----------------|
| LOCUS | BE148882 | 563 bp | mRNA | linear | EST 21-JUN-2000 |
| DEFINITION | CM2-HT0244-221199-038-f02 | HT0244 | Homo sapiens | CDNA, | mRNA sequence. |
| ACCESSION | BE148882 | | | | |

| KEYWORDS | EST. |
|----------|--------------|
| SOURCE | human. |
| ORGANISM | Homo sapiens |

REFERENCE
AUTHORS
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 563)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,

Locatio

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Jones, M.R.,

| TITLE | JOURNAL MEDLINE | COMMENT |
|--|--|----------|
| Shotgun sequencing of the human transcriptome with ORF expressed Sequence tags | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) | 20202663 |
| Goldman, G.H., Carvalho, A.F., Matsukuma, A., Batig, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J | | |

Tel.: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?lcr=et2-cm2-H70244-22>)
199-038.f026c3-1999-11-226t4-1)
Seq primer: puc_18 forward
High quality sequence start: 26
High quality sequence stop: 562.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Hr0244"
/dev_stage="Adult"
/node="Organ: head neck; Vector: puc18; Site:1: Smal;
Site:2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions.*

```

| | | | | |
|-----------------------|----------------|---|----------|------------|
| Query Match | 39.7% | Score 355 | DB 9 | Length 563 |
| Best Local Similarity | 100.0% | Pred. No. 5.7e-174 | | |
| Matches 355 | Conservative 0 | Mismatches 0 | Indels 0 | Gaps 0 |
| QY | 541 | gctctgcctttgttagagagctctgagcgccacctctccaatatgaacattctgaagcaaga | 600 | |
| | | | | |
| Db | 112 | gctctgcctttgttagagagctctgaagcccaactctccaatatgaacattctgaagcaaga | 171 | |
| | | | | |
| QY | 601 | agaaagatgagagacacactacagaacactctctctccacactactctccacatgtaacc | 660 | |
| | | | | |
| Db | 172 | AGAAAGAGAGACACACTACAGACACTCTTTCTCCACCACTCACTCCCACTGTACCC | 231 | |
| | | | | |
| QY | 661 | acccttaacatctccagctgctctcgaanaagaatgtttttaaagtaatttgtttgtg | 720 | |
| | | | | |
| Db | 232 | ACCCTTAATATATTCCTCACTGCTCTGAAAAACAAGTATTTTGAAGTCATTTTGTTGTG | 291 | |
| | | | | |
| QY | 721 | ctctctctagtgctctctctctcgtcaagtccttagacctgtgcccctccactaccagagct | 780 | |
| | | | | |
| Db | 292 | CTCTCTCTAGAGTCTCTCTCTCTCTGCTTAGGCGCTTGCCCTCCCTTACCAGAGCT | 351 | |
| | | | | |
| QY | 781 | tagagcttaattacctggaagaattccaggaagaactgtagctctccctagctagtgcatttaac | 840 | |
| | | | | |
| Db | 352 | TAGAGCTTAATTAACCTGGAAGAATTCAGAGAAACTGTAGCTTCCGTAGTGTCAATTATTAAC | 411 | |
| | | | | |
| QY | 841 | cttaaatgcaatcaaggaagaatgagcaacaagaagacaatataattttaaatgtgc | 895 | |
| | | | | |
| Db | 412 | CTTAATATGCATACGGAAGGTAGCAAAACGAAGTCATTAATTAATTTTAAATGTGC | 466 | |

| RESULT | 9 |
|------------|---|
| AI079465/c | 431 bp |
| LOCUS | mRNA |
| DEFINITION | linear |
| | EST 01-OCT-1998 |
| | 0239f05.s1 Soares_NhMMPu_S1 Homo sapiens CDNA clone IMAGE:1677729 |

| | |
|-----------|-----------------------|
| ACCESSION | 3', mRNA sequence. |
| VERSION | AI079465 |
| KEYWORDS | AI079465.1 GI:3415716 |
| SOURCE | EST. |
| ORGANISM | human. |
| | Homo sapiens |

REFERENCE (pages 1 to 431)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.

This clone is available royalty-free through LMD ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 921 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 431.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1b="IMAGE:1677729"
/clone_1lb="Scares_NhHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10b"
/note="Organ: mixed (see below); Vector: PT73D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NHM, pregnant uterus
NBHMPu, and fetal heart NBH19W) were mixed, and ss circi
was made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-488479."

```

| | Query Match | 37.9% | Score 339 | DB 9 | Length 431 | |
|----|-----------------------|---|--------------------|----------|------------|--|
| | Best Local Similarity | 100.0% | Pred. No. 1,2e-165 | | | |
| | Matches 339 | Conservative 0 | Mismatches 0 | Indels 0 | Gaps | |
| QY | 557 | agctctgagcccaactcttccaatlaaacaattctcaagccaagaagacagtggacaccc | 616 | | | |
| Db | 342 | AGCTCTGAGCGCCACCTCTTCCATTAATAACATCTCAGGCCAAGAAGACAGTGAGCACACC | 283 | | | |
| QY | 617 | taccagagacattcttcttctcccaactcaactctcccactgtacccaccacctaatctcc | 676 | | | |
| Db | 282 | TACCAAGACACTCTTCTTCCACACCTCAGCTCCCACTGTACCCACCCCTAAATCATTTCC | 223 | | | |
| QY | 677 | agtgctctcaaaaagaatgttttccaagaatcaattgtttgtgctctctcagttctt | 736 | | | |
| Db | 222 | AGTGCTCTCAAAAAGCATGTTTTTCAAGATCATTTTGTGTGTCCTCTTAGTGTCTT | 163 | | | |
| QY | 737 | ctctctctgctcagctcttaagcctgtgcccctcccttaaccagaagctttagcttaattacctg | 796 | | | |
| Db | 162 | CTTCTCTGCGCATCTTAAAGCTTGCCCTGCCCCTTACCACAGGCTTAGGCTTAATTAATCTG | 103 | | | |
| QY | 797 | aaagattccaaggaactgttagcttctcctagctagtgctcatttaaccttaaatgtcaatcag | 856 | | | |
| Db | 102 | AAAGATTCCAGGAACGTAGTCTCTACTACTAGTCATTTAAACCTTAATGCAATCAGG | 43 | | | |
| QY | 857 | aaagtagcaaacgaagtcataataatattttttaaagtgc | 895 | | | |
| Db | 42 | AAAGTAGCAACAGAACTCAATBAATATTTTAAATGTC | 4 | | | |

| | | | |
|----------|-----------------------|---|-----------------------------|
| | RESULT | 10 | |
| B1770944 | LOCUS | B1770944 | 820 bp mRNA EST 25-SEP-2001 |
| | DEFINITION | 603059769F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5209073 5' , mRNA sequence. | |
| | ACCESSION | B1770944 | |
| | VERSION | B1770944.1 | GI:15762522 |
| | KEYWORDS | EST. | |
| | SOURCE | human. <i>Homo sapiens</i> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | |
| | ORGANISM | human. | |
| | REFERENCE | NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) | |
| | AUTHORS | Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs@email.nih.gov Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov | |
| | TITLE | JOURNAL | |
| | COMMENT | | |
| | FEATURES | | |
| | source | | |
| | | High quality sequence stop: 772. location/Dualifiers 1..820 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5209073" /clone_id="NIH_MGC_122" /lab_host="DH10B" /note="Organ: pooled lung and spleen; Vector: PCMV-SPORE6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dt primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 020. Note: This is a NIH/MGC library." | |
| | BASE COUNT | 225 a 237 c 162 g 196 t | |
| | ORIGIN | | |
| | | | |
| | Query Match | 37.3%; Score 334; DB 10; Length 820; | |
| | Best Local Similarity | 100.0%; Pred. No. 5.7e-163; | |
| | Matches 334: | Conservative 0; Mismatches 0; Indels 0; Gaps 0; | |
| OY | 378 | aagaaattctgacagtgtctggctgcgaagaagcaatgcccttgatcatltaa | 437 |
| Dd | 270 | AAGAAAATTTCATGACAGTGCTGGGCTCCCAAAGAACAAGCAGTGCCCTCGATCATTTCAA | 329 |
| OY | 438 | ggcgcaatgtgaagaanaaacacaacaaaaggcacacagaaagccaacaaacgatccag | 497 |
| Dd | 330 | GGGCAATGTGAAGAAACACACACCANAAGCACACAGAACCAACAAACACTTCCAG | 389 |
| OY | 498 | agacctgcagcaattttcaaacaatgtcagctaagaagcttgtctgacttagta | 557 |
| Dd | 390 | AGCCTGCCAACAATTTCTCAAAACATGTACAGTAGAAGCTTGGCTGCTTTSTAGGA | 449 |
| OY | 558 | gctctgagcgcgcacacttcgaattaacattctcagcgaagaagaagttgascacc | 617 |
| Dd | 450 | GCTCTGACCGCCCACCTCTTCCAATTAACATTTCTAGCCAAAGAAACAGTGACACACT | 509 |
| OY | 618 | accagacactcttctcccacactcactctccacagtaccacacccccaaatatcca | 677 |
| Dd | 510 | ACCAAGACACTCTTCTTTCGCCACGCACTCTGCCACTGTACCCACCCCTAAATCATTTCCA | 569 |
| Oy, | 678 | gtgctctcaaaaaagcatgttlcttcaaatcattt | 711 |

[illegible]

[illegible][illegible]

```

FEATURES
source

Location/Qualifiers
1. 429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2732243"
/clone_1ib="NCI_CGAP_Sub6"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pRTT3D-Pac (Pharmacia) with a modified Sub6
polylinker. Site1: Not I; Site 2: Eco RI. NCI CGAP Sub6
is a subtracted library derived from BM, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BM was used as a tracer in a subtractive
hybridization with a driver comprising the IMAGE pool
(NCI_CGAP_Kid3 pool 1 LHAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LHAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 LHAM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 LHAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470993, 1475951-1476743
); NCI_CGAP_P122 pool 1 LHAM 2452-2459, 2758-2759,
3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959,
1217998-1220615); NCI_CGAP_C010 pool 1 LHAM 2644-2653,
2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351
). (50% of the driver population), plus a pool of 3,840

```

arrayed clones from NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of the driver population), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of the driver population). Subtraction was performed as previously described [Bonaldo & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. *Genome Research* 6, 791-806].

TAG_L1B=NCI_CGAP-Lu13
TAG_T1SSU=Lung
TAG_SEQ=GCCGG*

92 a 82 c 112 g 140 t 3 others

| Query Match | Similarity | 35.2% | Score 315; | DB 9; | Length 429; |
|-------------|------------|--|-----------------|------------|-------------|
| Best Local | Similarity | 99.7%; | Pred. 4,1e-153; | | |
| Matches | 365; | Conservative | 0; | Mismatches | 1; |
| | | | | Indels | 0; |
| | | | | Gaps | 0; |
| QY | 358 | tggttcctcgagagcccccagagaagaanaatlccatgacagtgctcgtggcctgcaccaagaagcag | 417 | | |
| DB | 383 | tggTTCCTGAGAGCCCGGAGAAATTCATGACAGTGTCTGGGGCTGCCAAGAAGCAG | 324 | | |
| QY | 418 | tgccccctgatacatlccaagtgcaatgtgaaagaanaagccaccaaagagacacagaa | 477 | | |
| DB | 323 | TGCCCCCTGATCATTTCAAGGGGCAATGTGAAGAAACAGACACCAAGGACACACAGA | 264 | | |
| QY | 478 | aagccaaacaagcattlccagagcctgcgcagcaattlctcaaaacaatlytcagctaaagac | 537 | | |
| DB | 263 | AAGCCAAACAACATTTCCAGAGCTGCGCAGCAATTTCTCAAAACATGTGCAGCTAAGAAC | 204 | | |
| QY | 538 | ttgtctctgacctttagagagctctcgagcgcccaactcttccaattaaacattctcaagca | 597 | | |
| DB | 203 | TTTGTCTGTGCCCTTTGTAGAGGCTGTGAGGGCCCACTCTTCATTTAAATTTCTACGCCA | 144 | | |
| QY | 598 | agaagacagtgagacaacactaccagaacactctctctccacactcaactctcccaactgta | 657 | | |
| DB | 143 | AGAAGACAGTGAAGCACACNCTTACCAAGACACTCTTCTCTCCACACTCAGTCCCACTGTA | 84 | | |
| QY | 658 | cccaacccctaaatcaltccactgactctcctcaaaaagcatttcttaagaatcatttggttg | 717 | | |
| DB | 83 | CCCAACCCCTAAATCATTTCCAGTGCTCTCAAAAAGCATGTTTTTCAAGATCATTTTGTGTTG | 24 | | |
| QY | 718 | ttgtctc 723 | | | |
| DB | 23 | TTGTCTC 18 | | | |

| RESULT 13 | LOCUS AA317141.C | DEFINITION EST119048 Lung Homo sapiens cDNA 5' end, mRNA | EST 19-APR-1997 |
|-----------|---|--|-----------------|
| VERSION | AA317141 | | |
| KEYWORDS | AA317141.1 GI:1969489 | | |
| EST. | | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 297) | | |
| | Adams, M.D., Keriavaga, A.R., Fleischmann, R.D., Fulcher, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Well, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Eale-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodek, A., Gnellm, C.L., Hanna, M.C., Heblum, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Lau, H.-I., Marrairos, S.M., Merrick, J.M., Moreno-Palancques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, U.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Yi, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dinke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A., He, W.W. | | |

TITLE Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Frazer, C.M. and Venter, J.C.
JOURNAL Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)
COMMENT 96026280
 Other ESTs: THC160246
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi.html>)
 Seq primer: M3 Reverse

FEATURES

source Location/Qualifiers
 1..297
 /organism="Homo sapiens"
 /db_xref="ATCC (Inhost):117611"
 /db_xref="taxon:9606"
 /clone_lib="Lung"
 /dev_stage="adult"
 /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI ; Site_2: XhoI"
BASE COUNT 88 a 43 c 86 g 80 t
ORIGIN

Query Match 33.2%; Score 297; DB 9; Length 297;
Best Local Similarity 100.0%; Pred. No. 9.3e-144;
Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

587 attctcagcagaagaagtagtagcacacacctccacagacactctctcccccact 646
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 297 ATTCTTAGCCAGAGACAGTGCACACCTTACACACACTCTTCTTCCACCTCCTACT 238
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 647 ctccaccatgtaccacacccaaatcattccagtgctctcaaaaagcatgttttcaagt 706
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 237 CTCGCCAGTGTACCCACCCCTTAATCATTCACAGTGCCTCAAAAAGCATGTTTTCAGAT 178
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 707 caattgttctgtctctctctagtgctctctctctctctcagcttagcctgtgcctc 766
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 177 CATTTTGTGTGTGCTCTCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 118
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 767 cccttaaccagcttaggttaattaccggaagatccaggaactgttagctctctagc 826
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 117 CCTTACCCAGGCTTAGGCTTAATTACCTGAAGATTCAGGAACTGTAGCTCTCTAGC 58
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 827 tagtgatcattaaaccttaaatgcaatcaggaagtagcaaacaggaagtaataata 883
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 57 TACTGTCAATTAACTTAATGCAATCAGGAAGTAGCAAAAGTAGCAATTAATAA 1

RESULT 14
LOCUS AI830551 491 bp mRNA linear EST 21-DEC-1999
DEFINITION wj51b02.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2406315 3', mRNA sequence.
ACCESSION AI830551
VERSION AI830551.1 GI:5451222
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 491)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/nclogap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: www.bio.lnlnl.gov/bbrp/image/image.html
 Insert length: 406 Std Error: 0.00
 Seq primer: -40up from Gibco
 High quality sequence stop: 463.

FEATURES

source Location/Qualifiers
 1..491
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:2406315"
 /clone_lib="NCI_CGAP_Lu19"
 /tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pTZ19-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTZ19 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 129 a 81 c 132 g 145 t 4 others
ORIGIN

Query Match 30.1%; Score 269; DB 9; Length 491;
Best Local Similarity 100.0%; Pred. No. 4.3e-129;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

626 ctcttcttccacactcactccacatgaccaccccttaataatccagtgctc 665
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 269 CTCTTCTTCCACCTCACTCTCCACCTTACACCCCTTAATATTCACAGTCTCTC 210
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 686 aaaaagcatgttttcaagatcatttggttgtgctctctcagtgctctctctcg 745
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 209 AAAAAGCATGTTTTCAGATCATTTTGTGTGCTCTCTCTCTCTCTCTCTCTCT 150
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 746 tcagtttagcctgtgcccctccctaccagagcttaggcttaatactgaaagatcc 805
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 149 TCAGTCTTAGCCCTGTGCCCTTACCAGAGCTTAGGCTTAATTAACCTGAAGAATCC 90
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 806 aggaacatgtagcttccctagctagtgatcaattcaacttaatgcaatcaggaagtagca 865
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 89 AGGAACATGTAGCTTCTCTAGCTAGTGTCAATTAACTTAATGCAATCAGGAAGTAGCA 30
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 866 aacagagatcaataatattttaatgt 894
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 29 AACGAAAGTCATTAATATTATTTTAAATGT 1

RESULT 15
LOCUS AI860310 270 bp mRNA linear EST 26-AUG-1999
DEFINITION w101e10.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2423658 3', mRNA sequence.
ACCESSION AI860310
VERSION AI860310.1 GI:5513926
KEYWORDS EST.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 01:42:48 ; Search time 3013.46 Seconds
(without alignments)
4008.606 Million cell updates/sec

Title: US-09-700-770-5
Perfect score: 895
Sequence: 1 ctaatcgtctacgacgc.....aataatattttaatgac 895

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estb1:*
2: em_estb2:*
3: em_estb3:*
4: em_estb4:*
5: em_estb5:*
6: em_estb6:*
7: em_estb7:*
8: em_estb8:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_plo:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 672.2 | 75.1 | 820 | 10 | BT770944 603059769 |
| 2 | 582.6 | 65.1 | 613 | 10 | BG536626 60256231 |
| 3 | 510.6 | 57.1 | 667 | 9 | AW471176 xvi3a03.x |
| 4 | 497 | 55.5 | 498 | 10 | BF196945 |
| 5 | 488.4 | 54.6 | 491 | 9 | AI340056 qn18h09.x |
| 6 | 487.2 | 54.4 | 504 | 9 | AI809681 w76c06.x |
| 7 | 471 | 52.6 | 491 | 9 | AI830551 w51b07.x |
| 8 | 465.2 | 52.0 | 482 | 10 | BF061539 7j52910.x |
| 9 | 464 | 51.8 | 477 | 9 | AI142875 ox64e06.s |
| 10 | 431.8 | 48.2 | 461 | 9 | AI1276107 q174h09.x |
| 11 | 427.6 | 47.8 | 563 | 9 | BE148882 CM2-HT024 |
| 12 | 426 | 47.6 | 426 | 9 | AA479746 zuz3g03.s |
| 13 | 414.8 | 46.3 | 419 | 9 | AA430487 zw23f06.s |
| 14 | 409 | 45.7 | 429 | 9 | AM298444 UT-H-BW0 |
| 15 | 405 | 45.3 | 431 | 9 | AI079465 oz39f05.s |
| 16 | 397 | 44.4 | 397 | 9 | AA481852 zv42f04.r |
| 17 | 375.2 | 41.9 | 404 | 10 | BF745581 QVO-BT084 |

| | | | | | |
|----|-------|------|------|----|--------------------|
| 18 | 365 | 40.8 | 392 | 9 | AA433968 zw23f07.r |
| 19 | 341.2 | 38.1 | 356 | 9 | AM770737 h189g02.x |
| 20 | 339.4 | 37.9 | 409 | 9 | BE073132 MR3-BT055 |
| 21 | 317.4 | 35.5 | 410 | 10 | BF745586 QVO-BT084 |
| 22 | 305.6 | 34.1 | 356 | 9 | AM173795 xp96910.x |
| 23 | 303 | 33.9 | 422 | 10 | BF743581 |
| 24 | 301.2 | 33.7 | 326 | 10 | BF882906 QVO-BT084 |
| 25 | 297 | 33.2 | 297 | 9 | AA317141 EST19048 |
| 26 | 287.6 | 32.1 | 302 | 9 | AM770575 |
| 27 | 270.4 | 30.2 | 408 | 9 | AM183010 |
| 28 | 269.4 | 30.1 | 387 | 10 | BM430624 |
| 29 | 269 | 30.1 | 275 | 9 | AA366288 EST77218 |
| 30 | 267 | 29.8 | 270 | 9 | AI860310 w101e10.x |
| 31 | 264.8 | 29.6 | 411 | 9 | AI368232 qw16e02.x |
| 32 | 263 | 29.4 | 263 | 10 | BE645174 7e64d07.x |
| 33 | 263 | 29.4 | 263 | 10 | BF055365 |
| 34 | 258.4 | 29.0 | 479 | 10 | BE624584 ud3n11.y |
| 35 | 258.6 | 28.9 | 605 | 9 | AA537274 vk46b02.r |
| 36 | 257.8 | 28.8 | 466 | 9 | AA881969 vx31g04.r |
| 37 | 257.8 | 28.8 | 484 | 9 | AA789593 vw61f07.r |
| 38 | 254 | 28.4 | 255 | 9 | AI207071 qf80f02.x |
| 39 | 253.4 | 28.3 | 433 | 9 | BE197231 ug73b06.y |
| 40 | 241.2 | 26.9 | 271 | 9 | BE181362 CM0-HT063 |
| 41 | 240.4 | 26.9 | 829 | 10 | BI078998 602873258 |
| 42 | 240.4 | 26.9 | 1279 | 10 | BG535485 602563170 |
| 43 | 238 | 26.6 | 242 | 9 | AA477092 zu35g03.r |
| 44 | 234 | 26.1 | 251 | 10 | BF743524 QVO-BT084 |
| 45 | 219 | 24.5 | 254 | 9 | AM389442 ILO-ST016 |

ALIGNMENTS

RESULT 1
LOCUS BT770944 820 bp mRNA linear EST 25-SEP-2001
DEFINITION 603059769P1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5209073 5',
mRNA sequence.
ACCESSION BT770944
VERSION BT770944.1 GI:15762522
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 820)
NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strusberg, Ph.D.
Email: cga@pds-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LAM11524 row: j column: 18
High quality sequence stop: 772.
Location/Qualifiers
1..820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5209073"
/clone_id="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size

FEATURES

source

range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC library."

BASE COUNT 225 a 237 c 162 g 196 t

ORIGIN

Query Match 75.1%; Score 672.2; DB 10; Length 820;
Best Local Similarity 98.4%; Pred. No. 2,8e-163;
Matches 721; Conservative 0; Mismatches 8; Indels 4; Gaps 4;

QY 108 cctgagatcctcccaaggatcgaagagcctcgaaggaacccctcatalatctcaca 167
Db 1 cctgagatcctcccaaggatcgaagagcctcgaaggaacccctcatalatctcaca 59
QY 168 gcaactaagagctgcaccgagcagtgatgataaagttctatctctcctcctcgt 227
Db 60 gcaactaagagctgcaccgagcagtgatgataaagttctatctctcctcctcgt 119
QY 228 gctgcacaatagtctgatagtctcctcgaagcctcgaagggctgcgaagag 287
Db 120 gctgcacaatagtctgatagtctcctcgaagcctcgaagggctgcgaagag 178
QY 288 ccaacagagagcagagcgaagcctcgaagagagtgctcgaagagcgaagagtg 347
Db 179 ccaacagagagcagagcgaagcctcgaagagagtgctcgaagagcgaagagtg 238
QY 348 gtgcaaaagtgttctctg-agagcccgagaaagaatcagaagtgctcggctgc 406
Db 239 gtgcaaaagtgttctctcctcgaagcctcgaagagagtgctcggctgc 298
QY 407 caaagaagcagtgccctgtgatactcgaagagcagtggaagaacaagaacacaa 466
Db 299 caaagaagcagtgccctgtgatactcgaagagcagtggaagaacaagaacacaa 358
QY 467 ggcacacagaagaagcgaagaacatccgaagcctgcgaagatctcgaagaatgc 526
Db 359 ggcacacagaagaagcgaagaacatccgaagcctgcgaagatctcgaagaatgc 418
QY 527 agctaagaagcttctcctcctcctcgaagagctcgaagcctcctcgaatgaac 586
Db 419 agctaagaagcttctcctcctcctcgaagagctcgaagcctcctcgaatgaac 478
QY 587 attctcagaagaagaagtgagacacactacagaacactctctcctcgaagcctc 646
Db 479 attctcagaagaagaagtgagacacactacagaacactctctcctcgaagcctc 538
QY 647 ctcccaacttaccaccccttaactcgaagctcgaagcctcgaagaatgtttcaagat 706
Db 539 ctcccaacttaccaccccttaactcgaagctcgaagcctcgaagaatgtttcaagat 598
QY 707 catctgtgtgtcctcctcctcgaagctcgaagcctcgaagaatgtttcaagat 766
Db 599 ca-cttgggttggtgctcctcgaagctcgaagcctcgaagaatgtttcaagat 657
QY 767 cccctacccaagcttagcttaactcgaagaatcgaagaatcgaagaatcgaagaat 826
Db 658 cccctacccaagcttagcttaactcgaagaatcgaagaatcgaagaatcgaagaat 717
QY 827 tagtgcatttaa 839
Db 718 tagtgcattttta 730

RESULT 2
BG536626 613 bp mRNA linear EST 03-APR-2001
LOCUS 602866231.1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4691118 5',
DEFINITION mRNA sequence.
ACCESSION BG536626
VERSION BG536626.1 GI:13528159
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 613)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE NIH-MGC http://imgc.ncbi.nlm.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLCM1509 row: m column: 07
High quality sequence stop: 611.
Location/Qualifiers
1. 613
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4691118"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1; Still (ggcgctcggc); Site: 2; Still (ggcattatggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGAGCGCGCATG-drr(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC library."

BASE COUNT 165 a 176 c 132 g 140 t

ORIGIN

Query Match 65.1%; Score 582.6; DB 10; Length 613;
Best Local Similarity 99.2%; Pred. No. 4.2e-140;
Matches 596; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 140 ggaagaaaccttccattatattctcgaagaacactcagctcagcagagtgagatga 199
Db 14 ggaagaaaccttccattatattctcgaagaacactcagctcagcagagtgagatga 73
QY 200 aagttctaactctcctcctcctcctcctcctcctcctcctcctcctcctcctc 259
Db 74 aagttctaactctcctcctcctcctcctcctcctcctcctcctcctcctcctc 133
QY 260 ggaagctgaatccagagtgctcgaagaagcgaagagcgaagcgaagcgaagcga 319
Db 134 ggaagctgaatccagagtgctcgaagaagcgaagagcgaagagcgaagcgaagcga 193
QY 320 ggcctcagaagagcgaagaatgtgagtgcaagaatgtgtcctcgaagcgaagcga 379
Db 194 ggcctcagaagagcgaagaatgtgagtgcaagaatgtgtcctcgaagcgaagcga 253
QY 380 gaaatcattagcagtgctcgtgcgaagaagcagtgccctgtgatacattcaag 439
Db 254 gaaatcattagcagtgctcgtgcgaagaagcagtgccctgtgatacattcaag 313
QY 440 gcaatgtgaagaagaagaacacacacacacacacacacacacacacacacacac 499
Db 314 gcaatgtgaagaagaagaacacacacacacacacacacacacacacacacacac 373
QY 500 cctgcagaacattctcgaagaatgtcgaagaatgtcgaagaatgtcgaagaatgtcga 559
Db 374 cctgcagaacattctcgaagaatgtcgaagaatgtcgaagaatgtcgaagaatgtcga 433
QY 560 tctgagcgcccaactctcgaagaatgtcgaagaatgtcgaagaatgtcgaagaatgtcga 619


```
|||||
Db 434 TCTGAGCGCCACTCTTCCAAATTAACATTCGACGCCAAGAGACAGTGCACACCTAC 493
Oy 620 cagaacatcttcttccacactcacttcaccatgtaaccccaatcattccagc 679
Db 494 CAGACACTCTTCTTCCACCTCACTCTCCACGTACCCACCCCTAATATATTCAGT 553
Oy 680 gctccaaaagaacatgcttttcaagatacttctgttctctctctctctctctt 739
Db 554 GCTTCAAAAGCATTGTTTTCAGATCATTTGCG-TCCGTCCTCTCTAGTCTCTT 612
Oy 740 c 740
Db 613 c 613

RESULT 3
AM471176 667 bp mRNA linear EST 24-PEB-2000
LOCUS xvi3a03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2812972 3', mRNA sequence.
ACCESSION AM471176
VERSION AM471176
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 667)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40bp from GIBCO
High quality sequence stop: 345.
Location/Qualifiers
source
1. 667
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2812972"
/clone_lib="Soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NDHL19W, testis NHT, and B-cell
NCL-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 176 a 127 c 186 g 174 t 4 others
ORIGIN

Query Match 57.1%; Score 510.6; DB 9; Length 667;
Best Local Similarity 90.4%; Pred. No. 1.8e-121;
Matches 565; Conservative 0; Mismatches 58; Indels 2; Gaps 2;
```

```
Oy 390 gacagtctggtggtcccaagaagcagtggtccctgtgatcatcttcaagggaatgtga 449
Db 505 GAGGtGtTtGgCtGcCtAACTAGcAGtGcCCCTGtGtCAATTCaAGGGAATGtGA 446
Oy 450 gaaacaagaacacccaagcaccacagaaagaacaaagcattccagagcctgcagca 509
Db 445 GAAACAAGACtCCAAAGGCACTACAGAAAGCAACGCAATTCACAGAGCCTGCAGCA 386
Oy 510 alttctaaacaatgctcagtaagaagcttgctctgccttgttagagctctgagcgc 569
Db 385 ATTCTCAACATATGTCAAGCTAAGAGCTTCCTGCTGCTGTGAGAGCTCTGAGCGC 326
Oy 570 cacttctcaatlaaacattccacagaagaacagatgagacacactcagacacact 629
Db 325 CACTCTCCCAATTAACATTCNCAGCCAGAAAGAAAGNAGCACACNACNACACTT 266
Oy 630 tcttctccacactcacttccacactgtaacccaccctaatcatctcagtgcttcaaa 689
Db 265 TCTTCTCCCACTCACTCTCCACGTACCCACCCCTAATCATTCAGTCTCAAAA 206
Oy 690 agcatgttttcaagaatattgttctgtctctctctctctctctctctctctcag 749
Db 205 ACCATGTTTTTCAAGATCATTTTGTGTTGTCCTCTAGTCTCTCTCTCTCTCTCT 146
Oy 750 tcttagctgtgctccctccctaccacagcttaagcttaacttaactgaaagattccaga 809
Db 145 TCTTAGCTGTGCTCCCTCCCTTACCAGGCTTAGGCTTAATTAATTCGAAAGATTCCAGA 86
Oy 810 aacttagcttctcagtagtgatcatttaaccttaaatgcaatcaggaagaatgacaaca 869
Db 85 AACTGTAGCTTCTTACGTAGTGTCTATTATTAATTCATTAATGCAATCAGAAAGTACCAACA 26
Oy 870 gaagtcataatattttaaatt 894
Db 25 GAAGTCATTAATTAATTAATTAATGT 1

RESULT 4
BF196945 498 bp mRNA linear EST 03-NOV-2000
LOCUS BF196945/C
DEFINITION IMAGE:3527932 3', mRNA sequence.
ACCESSION BF196945
VERSION BF196945.1 GI:11085527
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 498)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
High quality sequence stop: 487.
Location/Qualifiers
source
1. 498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3527932"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
```


| db | Accession | Version | Keywords | Organism | Source | Reference Authors Title | Journal Comment | Features |
|-------------------------|--|---|----------|--------------|-----------------|-------------------------|-----------------|----------|
| AI809681/c | AI809681 | | | Homo sapiens | | | | |
| LOCUS | wf76c06.x1 | 504 bp | mrna | linear | EST 07-JUL-1999 | | | |
| DEFINITION | IMAGE:2361514 3', mRNA sequence. | | | | | | | |
| ACCESSION | AI809681 | | | | | | | |
| VERSION | AI809681.1 | GI:5396247 | | | | | | |
| KEYWORDS | EST. | | | | | | | |
| SOURCE | human. | | | | | | | |
| ORGANISM | Homo sapiens | | | | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | | | | | | |
| AUTHORS | 1 (bases 1 to 504) | | | | | | | |
| TITLE | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . | | | | | | | |
| JOURNAL | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index | | | | | | | |
| COMMENT | Unpublished (1997) | | | | | | | |
| | Contact: Robert Strausberg, Ph.D. | | | | | | | |
| | Email: cgapdb-remail.nih.gov | | | | | | | |
| | This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. | | | | | | | |
| | Seq primer: -40UP from GIBCO | | | | | | | |
| | High quality sequence stop: 459. | | | | | | | |
| FEATURES | Location/Qualifiers | | | | | | | |
| source | 1..504 | | | | | | | |
| | /organism="Homo sapiens" | | | | | | | |
| | /db_xref="taxon:9606" | | | | | | | |
| | /clone_image="IMAGE:2361514" | | | | | | | |
| | /clone_lib="Soares_NFL_T_GRC_S1" | | | | | | | |
| | /lab_host="DH10B" | | | | | | | |
| | /note="Organ: pooled; Vector: pTY73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung MHL19W, testis NHT, and B-cell NCL-GAP-GCHI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo." | | | | | | | |
| BASE COUNT | 134 a 85 c 141 g 144 t | | | | | | | |
| ORIGIN | | | | | | | | |
| Query Match | 54.4%; Score 487.2; DB 9; Length 504; | | | | | | | |
| Matches | Local Similarity 99.2%; Pred. No. 2e-115; | | | | | | | |
| Match 500; Conservative | 0; Mismatches 3; Indels 1; Gaps 1; | | | | | | | |
| db | 392 | caagtctggagctggccaagaagagccctctgcatcatcttcaaggccaatgtgaaga | 451 | | | | | |
| | | | | | | | | |
| db | 504 | cagttcttgaggctg-cnaagagagcgtcccttgatcatcttcaaggccaatgtgaaga | 446 | | | | | |
| | | | | | | | | |
| db | 452 | aaacagaagacaccaaagagcaccacagaagaacaaacagcatctccagagctgcagcaat | 511 | | | | | |
| | | | | | | | | |
| db | 445 | aaacagaagacaccaaagagcaccacagaagaacaaacagcatcttccagagcttccacgacat | 386 | | | | | |
| | | | | | | | | |
| db | 512 | tcttccaacaatgctcagctaaagaagcttgcctctgctcttgtaaggagctctgagcgcca | 571 | | | | | |
| | | | | | | | | |
| db | 385 | ttttctcaaaacaaatgttcagctaaagaaacctttggctctcttttgtagagagctctgagcgcca | 326 | | | | | |
| | | | | | | | | |
| db | 572 | ctcttcacaataaactctcagcccaagaagaagtgagcaccactaccagaactcttc | 631 | | | | | |
| | | | | | | | | |
| db | 325 | ctcttcccaatttaaacattttcagcccaaaagacagtgaacacacctaccagacactcttgc | 266 | | | | | |
| | | | | | | | | |
| db | 632 | tcttcccaactcactctcccaactgaccacaccccaaaatcaatctcagtgctctcaaaaag | 691 | | | | | |
| | | | | | | | | |
| db | 265 | ttttctcccaactcactctcccaactgaccacaccccaaaatcaatctcagtgctctcaaaaag | 206 | | | | | |
| | | | | | | | | |
| db | 692 | catgtttttcaagaatcatcttggttggtgctctctagtgatctctctctctctgctcagtc | 751 | | | | | |
| | | | | | | | | |
| db | 205 | catgtttttttcaagaatcatcttggttggtgctctctagtgatctctctctctctgctcagtc | 146 | | | | | |
| | | | | | | | | |

| | | | |
|-----------------------|--|--|-------------------|
| 0y | 752 | ttagcccttgccctccctccctaccagcgcttggtcctaattaccggaagatccagagaa | 811 |
| Db | 145 | TTAGGCTGTGCCCTCCCTACCGAGGCTTATGAGCTTAATACCTGAAGAATTCACAGGAA | 86 |
| 0y | 812 | cctgctgcctccagtagtgcattcaattaaccttaaatgcaatccagaagaagtacaacaga | 871 |
| Db | 85 | CTGTAGCTTCCCTAGCTACTGTCATTTAACTTAATGCAATCAGCAGCAAGTACGAACAGA | 26 |
| 0y | 872 | agtcataataatttttaaatgctc | 895 |
| Db | 25 | AGTCATTAATATTTTAAATGTC | 2 |
| RESULT | 7 | | |
| AI830551/c | | 491 bp | mRNA |
| LOCUS | | | linear |
| DEFINITION | wj51b02.x1 NC1_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2406315 | 3' | |
| ACCESSION | AI830551 | | |
| VERSION | AI830551.1 | GI:5451222 | |
| KEYWORDS | mRNA sequence. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo. | | |
| TITLE | 1 (bases 1 to 491) | | |
| | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. | | |
| | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), | | |
| | Tumor Gene Index | | |
| | Unpublished (1997) | | |
| JOURNAL | Contact: Robert Strausberg, Ph.D. | | |
| COMMENT | Email: cgapbs-remail.nih.gov | | |
| | Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. | | |
| | Emmert-Buck, M.D., Ph.D. | | |
| | cDNA Library Preparation: M. Bento Soares, Ph.D. | | |
| | cDNA Library Arrayed by: Greg Lennon, Ph.D. | | |
| | DNA Sequencing by: Washington University Genome Sequencing Center | | |
| | Clone distribution: NCI-CGAP clone distribution information can be | | |
| | found through the I.M.A.G.E. Consortium/LLNL at: | | |
| | www-bio.llnl.gov/dbrrp/image/image.html | | |
| | Insert length: 406 Std Error: 0.00 | | |
| | Seq primer: -40up from Gibco | | |
| | High quality sequence stop: 463. | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..491 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /clone="IMAGE:2406315" | | |
| | /clone_lib="NCI-CGAP_Lu19" | | |
| | /tissue_type="squamous cell carcinoma, poorly | | |
| | /differentiated (4 pooled tumors, including primary and | | |
| | metastatic)" | | |
| | /dev_stage="adult" | | |
| | /lab_host="DH10B (phage-resistant)" | | |
| | /note="Organ: lung; Vector: pT73B-Pac (Pharmacia) with a | | |
| | modified polylinker; 1st strand cDNA was prepared from | | |
| | pooled lung tumor tissue, and was then primed with a Not I | | |
| | : oligo(dT) primer. Double-stranded cDNA was ligated to | | |
| | Eco RI adaptors (Pharmacia), digested with Not I and | | |
| | cloned into the Not I and Eco RI sites of the modified | | |
| | pT73 vector. Library went through one round of | | |
| | normalization. Library constructed by Bento Soares and M. | | |
| | Fatima Bonaldo." | | |
| BASE COUNT | 129 a | 81 c | 132 g |
| ORIGIN | | 145 t | 4 others |
| Query Match | 52.6% | Score 471; | DB 9; Length 491; |
| Best Local Similarity | 97.1% | Prod. NO. 3.1e-111; | |
| Matches | 477; Conservative | 14; Indels | 0; Gaps 0; |
| 0y | 404 | taccgaagaagacgtgccctctatgctttcaaacgccaatgtagaagaacagagacc | 463 |

Db 491 TGGCAAGAGCACTGGCCCTGTGATCATTTCAAGGGCAATGTGAAGAAAACAAGACACC 432
 Qy 464 aagagccacacagaagcacaacagatccagagcctccagcaattctcacaacat 523
 Db 431 AAGAGCACACAGAAAGCAACAGATTCAGAGCCCTGCGACCAATTTCTCAACAAAT 372
 Qy 524 gtcagctaaagaagcttgcctcgtcttctgtagagctctgagcgcacactctccaat 583
 Db 371 GTGAGCTAAGAAGCTTGTCTGCTGTGTAGAGCTCTGAGGGCCCACTTTCCAAATTA 312
 Qy 584 aacattctcagcagaagaagcagtgagcacactaccagaacactctcttctccaccc 643
 Db 311 AACCTTCMCAGCCAGAGAGAGAGAGACANNACAGANCTCTCTCCACCTC 252
 Qy 644 actctccacatgtaccacccccaatcatctccagctctcctcaaaaagcatttttcaa 703
 Db 251 ACCTGCCACATGTACACACCCCTAAATCATTCAGTCTCTCAAAAAGCATGTTTTTCA 192
 Qy 704 gatacatttctgttctctctctctagctctctctctctctctctctctctctctctg 763
 Db 191 GATCATTTTGTCTGTCT 132
 Qy 764 ctccacctaccagccttaagcttaatactaccgaagaatccaggaactctgtagctct 823
 Db 131 CTCCTCTTACCCAGGCTTGAAGCTTAATTAACCTGAAGATTCAGGAACCTGTAGCTTCT 72
 Qy 824 agctagctcatttaacctaataatgcaatcaggaagaagtagcaacaagaatcaataa 883
 Db 71 AGCTAGGTGATTTAATCAATTAATGCAATGAGAAAGTACGAACAGAGAGTCAATTAATA 12
 Qy 884 tttttaaatgt 894
 Db 11 TTTTAAATCT 1

RESULT 8
 BF061539 482 bp mRNA linear EST 16-OCT-2000
 LOCUS BF061539.1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:3390114 3', mRNA sequence.
 ACCESSION BF061539
 VERSION BF061539.1 GI:10820370
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 482)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LINL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -400p from Gihco
 High quality sequence stop: 470.
 Location/Qualifiers
 1..482
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3390114"
 /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and ss circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from

the same 5 libraries. The pools consisted of the following
 libraries and clones: Soares NB2HP pool 1:
 309384-310919, 323208-325895 Soares NB2HP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HP-9W pool 1:
 758280-760583, 772104-774407 Soares NBHPA pool 1:
 304776-306311, 320136-322823, 326280-326663 Soares NBHP
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 131 a 77 c 134 g 140 t
 ORIGIN

Query Match 52.0%; Score 465.2; DB 10; Length 482;
 Best Local Similarity 99.2%; Pred No. 9.7e-110;
 Matches 478; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 415 cagtgccctgtatcatcattcaagggaatgtgaagaacaacagacacagcacc 474
 Db 482 CAGTGCCCTGTGATCATTTCAAGGGCAATGTGAAGAAAACAAGACACCAAGGACCCAC 423
 Qy 475 agaaagcacaacagaagatccagagcctgcagcaattctcaacaatgtcagtaaga 534
 Db 422 AGAAAGCACAAGCAAGCATTCAGAGCCTGCAGCAATTTCTCAACCAATGTACAGTAAGA 363
 Qy 535 agcttgcctcgtctgtgtagagctctgagcgcacactctcaatlaaacattctgag 594
 Db 362 AGCTTGTCTGTGCTTTGTAAGAGCTCTAGCCGCCACCTTTCATTAATTAATTCGCCAG 303
 Qy 595 ccaagaagaagtgagcacact-accagacactctctctccacactctccac 653
 Db 302 CCAAGAAGACAGTGAAGCACACCTTAACAGACACTTTTCTTCCACCTCACTGCCAC 243
 Qy 654 tttaccaccccaataatcattccagtgctctcaaaaagatgttttcaagaattttg 713
 Db 242 TGTACCCACCCCTAATATTCAGTCTCTCAAAAAGCATTTTTCAGATCAATTTTG 183
 Qy 714 ttgtgtcctctcagtgctctctctctctctcagctcagctcgtgcccctccatc 773
 Db 182 TTTGTTGCTCTCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 123
 Qy 774 ccagcctgaagcttaattcactcgaagaagatccagaagaactgtagctctcagtagtgc 833
 Db 122 CCAGGCTTGAAGCTTAATTAATTCAGAAAGATTCAGAAACTGTAGCTTCTACTAGTGC 63
 Qy 834 attaaccttaatacgaatcaggaagaagtagcaagaagaagtcataatattttaagt 893
 Db 62 ATTTAACCTTAATCAATCAAGAAAGTAGCAACAGAGATCAATTAATTTTAAATG 3
 Qy 894 tc 895
 Db 2 TC 1

RESULT 9
 A1142875/c 477 bp mRNA linear EST 23-OCT-1998
 LOCUS A1142875
 DEFINITION ox64e06.s1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1661122
 3', mRNA sequence.
 ACCESSION A1142875
 VERSION A1142875.1 GI:3659234
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 477)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 889 Std Error: 0.00
Seq primer: -40ml3 fwd. ER from Amersham
High quality sequence stop: 383.

FEATURES

source

1. .477
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="IMAGE:1661122"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: p7773D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 129 a 74 c 135 g 139 t
ORIGIN

Query Match 51.8%; Score 464; DB 9; Length 477;
Best Local Similarity 99.8%; Pred. No. 2e-109;
Matches 475; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy 420 cccctgtatatttcaaggagtgtaagaacaagaccacccaaggcaccacagaa 479
|||||
Db 477 cccctgtatatttcaaggagtgtaagaacaagaccacccaaggcaccacagaa 418
Oy 480 gccaaacaagcattccagaagcctgccaagaatttccaacatgtccaagaagctt 539
|||||
Db 417 gccaaacaagcattccagaagcctgccaagaatttccaacatgtccaagaagctt 358
Oy 540 tgcctgcctttgttagagagcttgagcgccacattcccaatlaaacattccaagc 599
|||||
Db 357 t-cttgccttggtagagagcttgagcgccacattcccaatlaaacattccaagc 299
Oy 600 aagaagagagagacacccatccagaacatttcttccacattacttccacatgtaac 659
|||||
Db 298 aagaagagagagacacccatccagaacatttcttccacattacttccacatgtaac 239
Oy 660 caccctaatcatcctcagtgctccaagaacatgttttcaagatcatattgtttgt 719
|||||
Db 238 caccctaatcatcctcagtgctccaagaacatgttttcaagatcatattgtttgt 179
Oy 720 gctctctagtgcttcttctcgtcagctttagccttgccctccatccagagc 779
|||||
Db 178 gctctctagtgcttcttctcgtcagctttagccttgccctccatccagagc 119
Oy 780 tttagcttaattacatgaagaatccagaacatgtagcttcttagtagtgtaatttca 839
|||||
Db 118 tttagcttaattacatgaagaatccagaacatgtagcttcttagtagtgtaatttca 59
Oy 840 ccttaaatgcaatccagaagtagcaaacagaatcaataatttttaaatgttc 895
|||||
Db 58 ccttaaatgcaatccagaagtagcaaacagaatcaataatttttaaatgttc 3

RESULT 10
AI276107/c 461 bp mRNA linear EST 29-JAN-1999
LOCUS AI276107
DEFINITION g114h09.x1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1878113
ACCESSION AI276107
VERSION AI276107.1 GI:3898381

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 461)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 412 Std Error: 0.00
Seq primer: -40UP from GIBCO.

FEATURES

source

1. .461
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="IMAGE:1878113"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: p7773D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 122 a 74 c 128 g 136 t
ORIGIN

Query Match 48.2%; Score 431.8; DB 9; Length 461;
Best Local Similarity 98.9%; Pred. No. 4.3e-101;
Matches 455; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Oy 436 aaggagcaatgtgaagaaacaaagacccacagaagaagcacaagcattcc 495
|||||
Db 461 aaggagcaatgtgaagaaacaaagacccacagaagaagcacaagcattcc 402
Oy 496 agagcctgcaagaatttctcaaacatgtcagctaagaagctttgctgctttag 555
|||||
Db 401 agagcctgcaagaatttctcaaacatgtcagctaagaagctttgctgctttag 342
Oy 556 gagctctgagggccacatttccaattaaacattcagccaagaagacagtgaacac 615
|||||
Db 341 gagctctgagggccacatttccaattaaacattcagccaagaagacagtgaacac 281
Oy 616 ctacccaagacatcttcttccacactacttccacagtgtccaccccttaattac 675
|||||
Db 282 ctacccaagacatcttcttccacactacttccacagtgtgtccaccccttaattac 224
Oy 676 cagtgctctcaaaagacatgttttcaagaatatttggtttgctcctcctagtgct 735
|||||
Db 223 cagtgctctcaaaagacatgttttcaagaatatttggtttgctcctcctagtgct 164
Oy 736 tcttctcgtcagcttagcctgtgacctcccttaaccagagcttaggcttaattact 795
|||||
Db 163 tcttctcgtcagcttagcctgtgacctcccttaaccagagcttaggcttaattact 104
Oy 796 gaaagattccagaagaatgtagcttcttagctagtgtagtcaatttaaccttaagcaatcag 855
|||||
Db 103 gaaagattccagaagaatgtagcttcttagctagtgtagtcaatttaaccttaagcaatcag 44
Oy 856 gaaagtagcaaacagaagcacaataataatttttaaatgttc 895

```

Db      43  GAAAGTAGCAACGAGTAGCATTAATTTTAAATGTC 4
|||||
RESULT  11
LOCUS    BE148882                      563 bp      mRNA      linear      EST 21-JUN-2000
DEFINITION CM2-HT0244-221199-038-f02 HT0244 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE148882
VERSION   BE148882.1 GI:8611606
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
          1 (bases 1 to 563)
          Dias Neto,E., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Nagai,M.A., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J
          Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
          Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
          20202663
          Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?el=ft2-CM2-HT0244-221
          199-038-f02&f3=1999-11-22&f4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 26
          High quality sequence stop: 562.
          Location/Qualifiers
            source          1..563
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /clone_lib="HT0244"
                           /dev_stage="Adult"
                           /note="Organ: head neck; Vector: puc18; Site.1: Smal;
                           Site.2: SmaI; A mini-library was made by cloning products
                           derived from ORESTES PCR (U.S. Letters Patent application
                           No. 196,716 - Ludwig Institute for Cancer Research)
                           profiles into the puc 18 vector. Reverse transcription of
                           tissue mRNA and cDNA amplification were performed under
                           low stringency conditions."
BASE COUNT      159 a      161 c      90 g      153 t
ORIGIN
Query Match      47.8%; Score 427.6; DB 9; Length 563;
Best Local Similarity 98.9%; Pred. No. 5.4e-100;
Matches 441; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
OY 450 gaacaagaacacccaagcaccacagaagaacaaacagcatctcagagctgcacaga 509
      |||||
Db 22 GAAATCAGACACCAAGGCA-CACAGAAAGCCAAACAGCATTCAGAGCCTGCACAGA 80
OY 510 atttccaagaatgtcagtaagaacttctgctgcttctgtgagagcttcgagcgc 569
      |||||
Db 81 ATCTCTCAACAGATGTCAGTAAGAGCTTCGCTCTGCTTTGTAGGAGACTCTGAGCGCC 140
OY 570 cactcttccaattaaacttctcagcacaagaacagatgagacacactcacaagactct 629
      |||||
Db 141 CACTCTTCCAAATTAACATTTCTCAGCCAAAGAGACAGTAGACACCTTACCAGACACTCT 200

```

```

OY 630 tcttcccaactcactcctccactgtaccaccaccctaaatcctcagtgctctcaaaa 689
      |||||
Db 201 TCTTCTCCACCTCAGCTCTCCACTGTACCCACCCCTAAATCATTCAGTCTCCAGAA 260
OY 690 agaatgttttcaagaatcattgttgttggctctctcctcagtgctctctctcagtg 749
      |||||
Db 261 AGCATGTTTTCAAGATCATTTTGTGTTGGCTCTCTAGTGTCTTCTCTCTCTCTCAG 320
OY 750 tcttagccttgctgcctccctaccaccagccttagcttaattaccctgaagaatccagga 809
      |||||
Db 321 TCTTAGCCTGTGCGCTCCCTTACCAGGCTTAGGCTTAATTACCTGAAAGATTCACAGA 380
OY 810 aactgtagcttctcagctagtgatcattcaacttaacctaataatcaggaagtagcaaca 869
      |||||
Db 381 AACTGTAGCTTCTTACTAGTGTCAATTTAACTTAATGCAATCAGAAAGTAGCAAAACA 440
OY 870 gaagtcataaatatttttaaatgtc 895
      |||||
Db 441 GAAAGTCAATTAATTTTAAATGTC 466

```

```

RESULT  12
LOCUS    AA479746                      426 bp      mRNA      linear      EST 08-AUG-1997
DEFINITION AA479746
          IMAGE:740020 3', mRNA sequence.
ACCESSION AA479746
VERSION   AA479746.1 GI:2205632
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
          1 (bases 1 to 426)
          Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
          Kucaba,T., Lacy,M., Le,N., Lennon,G., Merrit,M., Martin,J., Moore,B.,
          Schellenberg,K., Steptoe,M., Tan,F., Theisling,B., White,Y., Wyllie
          ,T., Waterston,R. and Wilson,R.
          WashU-Werck BST Project 1997
          Unpublished (1997)
          Contact: Wilson RK
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: estevatson.wustl.edu
          This clone is available royalty-free through LLNT; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Seq primer: -41m3 fwd. ET from Amersham
          High quality sequence stop: 419.
          Location/Qualifiers
            source          1..426
                           /organism="Homo sapiens"
                           /db_xref="GDB:5939727"
                           /db_xref="taxon:9606"
                           /clone="IMAGE:740020"
                           /clone_lib="Soares ovary tumor NbHOT"
                           /sex="female"
                           /tissue_type="ovarian tumor"
                           /lab_host="DH10B (ampicillin resistant)"
                           /note="Organ: ovary; Vector: pT73D (Pharmacia) with a
                           modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
                           strand cDNA was primed with a Not I - oligo(dT) primer [5'
                           TGTTACCAATCTGAAGTGGAGCGCGCGTCTTTTCTTTTCTTTTCTTTT
                           double-stranded cDNA was size selected, ligated to Eco RI
                           adapters (Pharmacia), digested with Not I and cloned into
                           the Not I and Eco RI sites of a modified pT73 vector
                           (Pharmacia). Library constructed by Bento Soares and
                           M. Relima Bonaldo."
BASE COUNT      121 a      64 c      122 g      119 t
ORIGIN

```

Query Match 47.6%; Score 426; DB 9; Length 426;
 Best Local Similarity 100.0%; Pred. No. 1.3e-99;
 Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 caccacagaagcgaacaaagcattccagagcctgcagcaattcttcaacaatgctcag 528
 |||||||
 Db 426 CACCACAGAAAGCAACAAACAGCATTCAGAGCCTGCCAGCAATTTCTCAACAAATGTGAC 367

QY 529 ctaagaagcttgctgccttgtagagctctgcagagcccaactcttccaattaacat 588
 |||||||
 Db 366 CTAAGAAGCCTTGGCTGCTGCTTTGTAGAGCTGTGAGGCCCCACTCTTCAATTAAACAT 307

QY 589 tctagcgaagaagacagctgagcagcaccctacagacactcttcttcccaactcact 648
 |||||||
 Db 306 TCTCAGCCAAAGACAGACAGTGCACACACTACAGACACTCTTCTTCCACCTCAGCTCT 247

QY 649 cccactgtaccacccccaataatcaltccagtgctctcaaaaagcatgttttcaagatca 708
 |||||||
 Db 246 CCCACTGTACCACCCCTTAATCATTCACAGTGCCTCAAAAAGCATGTTTTCAGATCA 187

QY 709 ttgtgtgtgtcctcctcctagtgctctctctcctcgtcaatcttagctgtgcccctcc 768
 |||||||
 Db 186 TTGTGTGTGTCTCTCTCTAGTGTCTTCTCTCTGTCAGTCTTAGGCTGTGCCCTCC 127

QY 769 ctacccagcttagagcttaattactcctgaagaattccaggaactgtagctctcctagcta 828
 |||||||
 Db 126 CTTACCCAGGCTTAGGCTTATTAACCTGAAGATTCAGGAAACGTGAGCTTCCAGCTA 67

QY 829 gtgtcatttaaccttaaatgcaatcagaagaagtagcaacagaagtcataataatttt 888
 |||||||
 Db 66 GTGTCAATTAACCTTAATGCAATCAGAGAGTAGCAACAGAGTCAATTAATTTT 7

QY 889 aaatgt 894
 |||||||
 Db 6 AAATGT 1

RESULT 13
 AA430487/c 419 bp mRNA linear EST 16-OCT-1997
 LOCUS 2x23106.s1 Soares ovary tumor NBH07 Homo sapiens cDNA clone
 DEFINITION IMAGE:770147 3', mRNA sequence.
 ACCESSION AA430487
 VERSION AA430487.1 GI:2111061
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 419)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Kucab,D., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
 Schellenberg,K., Stepien,M., Tan,F., Theising,B., White,Y., Wyllie,
 'T', Waterston,R. and Wilson,R.
 Washu-Merck EST Project 1997
 Unpublished (1997)
 CONTACT: Wilson RK
 ORGANISM Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -41ml3 fwd. 5' from Amersham
 High quality sequence stop: 356.
 location/Qualifiers
 1. 419
 /organism="Homo sapiens"
 /db_xref="GDB:596002"
 /db_xref="taxon:9606"
 /clone="IMAGE:770147"

/clone_1lb="Soares ovary tumor NBH07"
 /sex="Female"
 /tissue_type="Ovarian tumor"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Ovary; Vector: p773D (Pharmacia) with a
 modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAGTGGAGGCGCGGGGTGTTTTTTTTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified p773 vector
 (Pharmacia). Library constructed by Bento Soares and
 M.Falima Bonaído."

BASE COUNT 121 a 63 c 117 g 118 t
 ORIGIN

Query Match 46.3%; Score 414.8; DB 9; Length 419;
 Best Local Similarity 99.5%; Pred. No. 1.1e-96;
 Matches 416; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 478 aagcgaacaaagcattccagagcctgcagcaattctcaacaatgctcagtaagaagc 537
 |||||||
 Db 419 AAGCCAAACAAACATTCAGAGCCCTGCAGCAATTTCTCAACAAATGACGTAAGAAC 360

QY 538 ttgtctgccttgtagagcctgagagcccaactcttccaattaacattcagcca 597
 |||||||
 Db 359 TTGTGCTGCTTTTATAGAGCTGTGAGGCCCTCACTATTTCCATTAACATTTCTAGCCA 300

QY 598 agaagacagtagcacacactcagacactcttcttcccaactcactcctcactgta 657
 |||||||
 Db 299 AGAAGACAGTGCACACCTACCAACACTCTTCTTCCGCCACCTCAGCTCCACTGTA 240

QY 658 cccaccccaatcaatccagtgctctcaaaaagcatgttttcaagatcattgttg 717
 |||||||
 Db 239 CCCACCCCTTAATCAATTCAGAGCTCTCAAAAAGCATGTTTTCAGATCATTTGTTG 180

QY 718 ttgtctcctcagtgctctctctcctcgtcaatcttagcctgtgcccctccctccag 777
 |||||||
 Db 179 TTGCTCTCTAGTGTCTTCTTCTCTGTCAGTCTTAGGCTGTGCCCTCCATTACCCAG 120

QY 778 gcttagcttaattactcctgaagaattccaggaactgttagctcctagctagtgcatlt 837
 |||||||
 Db 119 GCTTAGGCTTATTAACCTGAAGATTCAGGAAACGTGATGCTTCTAGCTAGTGTGATTT 60

QY 838 aaaccttaaatgcaatcaggaagaagtagcaacagaagtcataataatcttaaatgtc 895
 |||||||
 Db 59 AACCTTAATGCAATCAGAGAAAGTAGCAACAGAGTCAATTAATTTTAAATGTC 2

RESULT 14
 AM298444/c 429 bp mRNA linear EST 16-JAN-2000
 LOCUS UI-H-BW0-a-1-e-06-0-UI.s1 NCI-CGAP_Sub6 Homo sapiens cDNA clone
 DEFINITION IMAGE:2732243 3', mRNA sequence.
 ACCESSION AM298444
 VERSION AM298444.1 GI:6705080
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 429)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA library preparation: M.B. Soares Lab Clone distribution:

NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LINL at:
www.bio.lnl.gov/bdrip/image/image.html
Seq primer: M13 Forward
POLY-A=yes

FEATURES

Source

Location/Qualifiers

1. 429

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2732243"

/clone_lib="NCI_CGAP_Sub6"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI CGAP_Brn50,
NCI CGAP_Lu13, NCI CGAP_Ov18, GBC1. The NCI CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI CGAP_Kid3 pool 1, LAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI CGAP_Kid5 pool 1
LAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI CGAP_Lu5 pool 1 LAM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI CGAP_GC4
pool 1 LAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983, 1475592-1476743
); NCI CGAP_Pr22 pool 1 LAM 2457-2459, 2758-2759,
3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959,
1217928-1220615); NCI CGAP_Co10 pool 1 LAM 2644-2653,
2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351
). (50% of the driver population), plus a pool of 3,840
arrayed clones from NCI CGAP_Sub1 (IMAGE Clonoids
2708616-2710535) and NCI CGAP_Sub2 (IMAGE Clonoids
2710536-2712455) (20% of the driver population), plus a
pool of 11,136 clones from NCI CGAP_Sub3 (IMAGE Clonoids
2712456-2723591) (30% of the driver population).
Subtraction was performed as previously described [Bonaldo
, Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806].

TAG_LIB=NCI_CGAP_Lu13

TAG_TISSUE=Lung

TAG_SEQ=GCCGG

BASE COUNT 92 a 82 c 112 g 140 t 3 others

ORIGIN

Query Match 45.7%; Score 409; DB 9; Length 429;
Best Local Similarity 99.3%; Pred. No. 3,4e-95;
Matches 409; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

312 taggagatgctccaggaagcgcccaagaatgtatgtaagatgttctgagagc 371
429 TAGGAGATGCTCCAGGAAGCGCCCAAGATGTGATGCAAGATGTTCTTGAGAGC 370
372 cccgaagaagaatcatcatcagctgtctggtgcgaagaagcagtcgctgtatca 431
369 CCCGAGGAAGAATTCATCAGACGTCTGCGGCCAAGAAGACGTCGCTGTGATCA 310
432 ttcaagaggaatgtgaagaagaacagagcaccacagaagaacccaagaagca 491
309 TTTCAAGGGCAATGTGAAGAAAAGACACCAAGGACCCACAGAAGGCCAACAAGCA 250
492 ttccagagcctgcccaggaatttctaaacaatgtcagctaaagaacttgcctctt 551
249 TTCCAGAGCTGCGCAGCAATTTCTCAACAAATGTCAGCTTAAGAACTTTGCTCTCCCTT 190
552 gtagagctctgaagcccaactcttcaaatiaaactctcagcagaagaagaagcagtgagc 611
189 GTAGAGCTCTGAGGCCACCTCTTCAATTAAACATTTCTCAGCCAAAGACAGTGAAC 130

QY 612 acacctaccagagacactcttctccacactcaactctccactgtaccacccctaarc 671
DB 129 ACACNTRACAGACACTCTTCTCTCCACCTCACTCCACAGTACCCACCCCTTAATC 70
QY 672 atccagctgctcccaagaacatgttttcaagatcaatttggctgtc 723
DB 69 ATTCCAGTGTCTCTCAAAAAGCATGTTTTCAGATCATTTTGTGTGCTC 18

RESULT 15
A1079465 431 bp mRNA linear EST 01-OCT-1998
LOCUS 0239f05.s1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1677729
DEFINITION 3', mRNA sequence.
ACCESSION A1079465
KEYWORDS A1079465.1 GI:3415716
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 431)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LINL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 921 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 431.
Location/Qualifiers

FEATURES

Source

1. 431

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1677729"

/clone_lib="Soares_NHMPu_S1"

/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"

/lab_host="DH10B"

/note="Organ: mixed (see below): Vector: pRT3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHMU, and fetal heart NbHM19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 121 a 65 c 124 g 121 t

ORIGIN

Query Match 45.3%; Score 405; DB 9; Length 431;
Best Local Similarity 99.5%; Pred. No. 3.7e-94;
Matches 427; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 468 goaccagagaagccaagaagcattccagagcctgcagcaatttccaacaatgtca 527
DB 431 GCACCAGAGAAAGCAACAGCATTCAGAG-CTGCCAGCAATTTCTCAACAATGTCA 373
QY 528 gctaaagacttgcctgtctgttagg-agctctgagcgcccaacttccaattaaac 586
DB 372 GCTAAGAGACTTTGCTGTGCTTTGTAGCAGCTCTGAGGCCACACTTCTCAATTAAC 313
QY 587 attctcagcagaagaagacagtgagacacactaccagacactcttctccactcaact 646
DB 312 ATTCTCAGCCAAAGAACAGTGAAGACACTTACAGACACTCTTCTCTCCACTCACT 253


```

Qy 647 ctcccactgtaccacccctaaatcattccagtgctctcaaaaagcatgtttttcaagat 706
    |||
Db 252 ctcccactgtaccacccctaaatcattccagtgctctcaaaaagcatgtttttcaagat 193
    |||
Qy 707 caatttgcttgctctctctctctctctctctctctctctctctctctctctctctc 766
    |||
Db 192 CATTTGTTGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 133
    |||
Qy 767 cccttaaccaggttaggttaattaccctgaagaattccaggaactgtagcttctctagc 826
    |||
Db 132 cccttaaccaggttaggttaattaccctgaagaattccaggaactgtagcttctctagc 73
    |||
Qy 827 tagtgatcattaaacctaaatgcgaatcagaagaagcaacagaagtcataaatattt 886
    |||
Db 72 TAGTGTCATTTAACCTTAATGCAATCAGGAAGTAGCAACAGAAAGTCAATTAATATT 13
    |||
Qy 887 ttaaatgtc 895
    |||
Db 12 TTAATGTC 4
    |||

```

Search completed: September 16, 2002, 01:42:54
 Job time: 3307 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:26:45 : Search time 89 Seconds
(without alignments)
2470.135 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895
Sequence: 1 ctaatctgtacgtacagc.....aataaatttttaaatgic 895

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/1na/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PTOUS_COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|-------------------|
| 1 | 56.8 | 6.3 | 7218 | 1 | US-08-232-463-14 |
| 2 | 38.6 | 4.3 | 7218 | 1 | US-08-232-463-14 |
| 3 | 38.6 | 3.8 | 289 | 4 | US-09-007-005-17 |
| 4 | 33.8 | 3.8 | 289 | 4 | US-09-244-796-17 |
| 5 | 32.8 | 3.7 | 445 | 4 | US-09-247-155-53 |
| 6 | 32.8 | 3.7 | 2686 | 4 | US-09-228-986-3 |
| 7 | 32.8 | 3.7 | 12001 | 1 | US-08-458-568A-11 |
| 8 | 32.2 | 3.6 | 6755 | 3 | US-08-931-999-4 |
| 9 | 31.6 | 3.5 | 1146 | 4 | US-08-596-684F-1 |
| 10 | 31.2 | 3.5 | 2264 | 4 | US-09-126-109-9 |
| 11 | 30.8 | 3.4 | 1698 | 1 | US-08-592-126-62 |
| 12 | 30.8 | 3.4 | 2062 | 1 | US-08-073-383-5 |
| 13 | 30.8 | 3.4 | 2062 | 3 | US-08-328-238A-3 |
| 14 | 30.8 | 3.4 | 2062 | 3 | PCT-US94-06365-5 |
| 15 | 30.8 | 3.4 | 2062 | 5 | PCT-US95-13661-3 |
| 16 | 30.8 | 3.4 | 13011 | 2 | US-08-791-849A-14 |
| 17 | 30.6 | 3.4 | 1311 | 4 | US-08-584-760A-66 |
| 18 | 30.6 | 3.4 | 7379 | 4 | US-09-341-587-5 |
| 19 | 30.4 | 3.4 | 4411529 | 4 | US-09-103-840A-1 |
| 20 | 30.2 | 3.4 | 7475 | 2 | US-08-971-036-1 |
| 21 | 30.2 | 3.4 | 7475 | 3 | US-09-096-570-1 |
| 22 | 30 | 3.4 | 277 | 4 | US-08-905-223-174 |
| 23 | 30 | 3.4 | 2218 | 2 | US-08-421-044-1 |
| 24 | 29.8 | 3.3 | 2606 | 4 | US-09-252-816A-7 |
| 25 | 29.8 | 3.3 | 3439 | 4 | US-09-252-816A-6 |
| 26 | 29.8 | 3.3 | 4168 | 3 | US-08-836-567-11 |
| 27 | 29.6 | 3.3 | 1100 | 2 | US-08-481-814A-5 |

| | | | | | |
|------|------|-----|------|---|-------------------|
| C 28 | 29.6 | 3.3 | 2323 | 4 | US-08-368-776A-6 |
| C 29 | 29.6 | 3.3 | 2323 | 5 | PCT-US96-00419-6 |
| C 30 | 29.6 | 3.3 | 2338 | 1 | US-08-425-069-1 |
| C 31 | 29.6 | 3.3 | 2338 | 2 | US-08-317-844B-1 |
| C 32 | 29.6 | 3.3 | 2456 | 1 | US-07-882-711-1 |
| C 33 | 29.6 | 3.3 | 2456 | 2 | US-08-462-174-1 |
| C 34 | 29.6 | 3.3 | 2457 | 3 | US-08-723-415B-12 |
| C 35 | 29.6 | 3.3 | 2457 | 2 | US-09-189-627A-12 |
| C 36 | 29.6 | 3.3 | 2492 | 1 | US-08-139-937-13 |
| C 37 | 29.6 | 3.3 | 2492 | 5 | PCT-US93-11310-13 |
| C 38 | 29.6 | 3.3 | 2517 | 1 | US-08-306-691B-18 |
| C 39 | 29.6 | 3.3 | 2517 | 2 | US-08-385-142-2 |
| C 40 | 29.6 | 3.3 | 2517 | 2 | US-08-481-814A-1 |
| C 41 | 29.6 | 3.3 | 2517 | 3 | US-08-801-092-2 |
| C 42 | 29.6 | 3.3 | 2517 | 4 | US-09-517-584A-3 |
| C 43 | 29.6 | 3.3 | 2901 | 4 | US-08-368-776A-4 |
| C 44 | 29.6 | 3.3 | 2901 | 5 | PCT-US96-00419-4 |
| C 45 | 29.6 | 3.3 | 4304 | 4 | US-08-368-776A-1 |

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F1s
; US-08-232-463-14

Sequence 6, Appl1
Sequence 6, Appl1
Sequence 1, Appl1
Sequence 1, Appl1
Sequence 1, Appl1
Sequence 1, Appl1
Sequence 12, Appl1
Sequence 12, Appl1
Sequence 13, Appl1
Sequence 13, Appl1
Sequence 18, Appl1
Sequence 2, Appl1
Sequence 1, Appl1
Sequence 2, Appl1
Sequence 3, Appl1
Sequence 4, Appl1
Sequence 4, Appl1
Sequence 1, Appl1

Query Match 6.3%, Score 56.8; DB 1; Length 7218;
Best Local Similarity 8.8%; Pred. No. 3.9e-08;
Matches 28; Conservative 170; Mismatches 122; Indels 0; Gaps 0;

QY 537 cttgctcgtccttctgtgagcctcgtgagcgcacactctcccaattcaattcagcc 596
Db 1171 YY 1230
QY 597 aagaagacagtgagcacactccagacactctctccacactcctccacagt 656
Db 1231 YY 1290
QY 657 accaccccaatcaatccagtcgtcctcaaaagcagtttcaagatcatttgtt 716
Db 1291 YY 1350
QY 717 gtgtcctcctcctagtcctctctctcctcgtcagcctgtgcccctccacca 776
Db 1351 YY 1410
QY 777 gacttaggcttaattaccctgaagatccaggaactgagctcctcagctagtgatc 836
Db 1411 YYYYYYYYYYYYYYYYYYYYYYYYYGGTACCAATTTCTATCTTTAACTTTCATA 1470
QY 837 taaccttaaatgcaatcag 856
Db 1471 GATAGTAATTCAGTATG 1490

RESULT 2
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 diagonal Road, Suite 500
City: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZpT-Fls
US-08-232-463-14

Query Match 4.3%, Score 38.6; DB 1; Length 7218;
Best Local Similarity 8.0%; Pred. No. 0.034;
Matches 20; Conservative 131; Mismatches 100; Indels 0; Gaps 0;

QY 273 agggctgcagcagggcagcagcagcagcagcagcagcagcagcagcagcagcagc 332
Db 1276 RRR 1217
QY 333 cggcacaagaatgtgagtgcaaatgtgtcctcagagcccgagaagaattcagatc 392
Db 1216 RRR 1157
QY 393 agtgcctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 452
Db 1156 RRR 1097
QY 453 aacaagacacacaaaggcagcagcagcagcagcagcagcagcagcagcagcagcagc 512
Db 1096 RRR 1037
QY 513 tctcaacaat 523
Db 1036 TCGGAATTAAT 1026

RESULT 3
US-09-007-005-17/c
Sequence 17, Application US/09007005B
Patent No. 6258558
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihc
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
FILE REFERENCE: 00786/350003
CURRENT APPLICATION NUMBER: US/09/007,005B
EARLIER FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064,491
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translation template
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(289)
OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 3.8%, Score 33.8; DB 4; Length 289;
Best Local Similarity 5.3%; Pred. No. 0.2;
Matches 11; Conservative 91; Mismatches 107; Indels 0; Gaps 0;

QY 624 cactctctcctccacactccacactgacacccctcaatcaattcagtgctc 683
Db 228 YAGYCYTYGYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNY 170
QY 684 tcaaaagcatgttttcaagatcattgtgtgtgtcctcctcagtgctcctctct 743

Db 169 YNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYS 110
Oy 744 cgtacgtttagcgcgtcccccctaccagcttaggttatactcgaagatt 803
Db 109 YNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYS 50
Oy 804 ccaggaactgtactcttagcttagt 832
Db 49 YGYTAAVATYTYGTYAAVATYATYATG 21

RESULT 4
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihue
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 3.8%; Score 33.8; DB 4; Length 289;
Best Local Similarity 5.3%; Pred. No. 0.2;
Matches 11; Conservative 91; Mismatches 107; Indels 0; Gaps 0;
Oy 624 cactctctctccaccctcactccaccgtaccacccccaataatcctcagtc 683
Db 229 YATGCTGTYGYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYS 170
Oy 684 tcaaaagacatgttttcaagatcattgtgtgtctctcctcagtcctctct 743
Db 169 YNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYS 110
Oy 744 cgtcagcttagcctgtccctccctaccagccttagcttaactgaagatt 803
Db 109 YNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYS 50
Oy 804 ccaggaactgtactcttagcttagt 832
Db 49 YGYTAAVATYTYGTYAAVATYATYATG 21

RESULT 5
US-09-247-155-53/c
; Sequence 53, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric

; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET 021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; EARLIER FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 53
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 81..356
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 81..152
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 6.2
; OTHER INFORMATION: seq ALLGSTWALTG/AL
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 406..411
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 429..445
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1
; OTHER INFORMATION: n=a, g, c or t
US-09-247-155-53

Query Match 3.7%; Score 32.8; DB 4; Length 445;
Best Local Similarity 53.0%; Pred. No. 0.54;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
Oy 207 aatctctctccctccctcctgtctgtccactaatgtatgtcattgtctcagaacct 266
Db 333 ACTCGGCTCGGGCTCTCTATCTGCTCGACGCTCGGGCTCGCACTCAT 274
Oy 267 gaatcagaggtcgcagagagccagagcagagcagagcagagcagagcagagc 326
Db 273 GAAAGTGGCCACAGATAGCCACAGTCCAGGAGCATAGACCGCGGACACACAGCA 214
Oy 327 ggaagcgagcca 338
Db 213 ACTAGCGGGCA 202

RESULT 6
US-09-228-986-3
; Sequence 3, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Neuenhulzen, Niels
; TITLE OF INVENTION: Compositions isolated from Plant Cells
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; EARLIER FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3

QY 439 ggcacatgtgaagaacacagaccacgaagca 470
Db 330 TCACACATGACATTACCAATAATCTCAAGGCA 361

RESULT 11

US-08-592-126-62/c
; Sequence 62, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Delinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TS2.seq
; US-08-592-126-62

Query Match 3.4%; Score 30.8; DB 1; Length 1698;
Best Local Similarity 54.4%; Pred. No. 5.2;
Matches 62; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 155 ttatatcttcacgaactacacagctgcacagatgctgcgatgaagtctatctctt 214
Db 1378 TTGGCTTCTTCACAGGAGGAGTCCGCTTCTCTCAGATGACATGAGAGGCTCTTGGCCCA 1319
QY 215 cccctccctgctgctgcacactaatgctgtagtcacatgctctcagagcctga 268
Db 1318 CCATCTTCCCTGTGCTGCCACGACTGCTGAACCGGATGTACGACAGGATCTTCA 1265

RESULT 12

US-08-073-383-5/c
; Sequence 5, Application US/08073383
; Patent No. 5443962
; GENERAL INFORMATION:
; APPLICANT: Draela, Giulio
; APPLICANT: Cottarel, Guillaume
; APPLICANT: Damaguez, Veronique
; TITLE OF INVENTION: [fillin] "Insert Title of Application" IASSAY AND REAGENTS FOR

; TITLE OF INVENTION: ANTI-PROLIFERATIVE AGENTS)
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHYE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/073,383
; FILING DATE: 19930604
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MIT-009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2062 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 211..1631
; US-08-073-383-5

Query Match 3.4%; Score 30.8; DB 1; Length 2062;
Best Local Similarity 58.9%; Pred. No. 5.8;
Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 673 ttccagtgctcacaagaatgttttcaagatcattgtgtgctgctcctagtg 732
Db 962 TTTCCTTGGCCAGAAAATAACTTTTCTTACTTTATCTGTAATTGTGCTTGAAT 903
QY 733 tctctctctcgtcagctttagcctgtgc 762
Db 902 TTTTCCACCGTCTCAGTGTGCGCTGTTC 873

RESULT 13

US-08-328-239A-3/c
; Sequence 3, Application US/08328239A
; Patent No. 6037136
; GENERAL INFORMATION:
; APPLICANT: Beach, David H.
; APPLICANT: Galationov, Konstantin
; APPLICANT: Jessus, Catherine
; TITLE OF INVENTION: Interactions between Raf Proto-Oncogenes
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLRY, HOAG & ELIOT
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:


```

;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(Text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328, 239A
;
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: CSW002.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 832-1000
; TELEFAX: (617) 832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2062 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 211..1631
; US-08-328-239A-3
;
;
; Query Match 3.4%; Score 30.8; DB 3; Length 2062;
; Best Local Similarity 58.9%; Pred. No. 5.8;
; Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
;
; QY 673 ttccagtgctctcaaaagcatgttttcaagatcatttggttgctctctctagtg 732
; || || || || || || || || || || || || || || || || || || || || ||
; DB 962 TTTCCTTGCCAGAAAAAATACCTTTTAACTTTATCTGTAATGTGTGCTTGAAAT 903
; || || || || || || || || || || || || || || || || || || || || ||
; QY 733 tctctctctcgtcagcttagcctgtgc 762
; || || || || || || || || || || || || || || || || || || || || ||
; DB 902 TTTCACCTGCTTCAGTCTTGCCCTGTTTC 873
; || || || || || || || || || || || || || || || || || || || || ||
;
; RESULT 14
; PCT-US94-06365-5/c
; Sequence 5, Application PC/TUS9406365
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay and Reagents for Identifying
; TITLE OF INVENTION: Anti-proliferative Agents
; NUMBER OF SEQUENCES: 6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06365
; FILING DATE: 06-JUN-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073, 383
; FILING DATE: 04-JUN-1993
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2062 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 211..1631
; PCT-US94-06365-5

```

```

;
; Query Match 3.4%; Score 30.8; DB 5; Length 2062;
; Best Local Similarity 58.9%; Pred. No. 5.8;
; Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
;
; QY 673 ttccagtgctctcaaaagcatgttttcaagatcatttggttgctctctctagtg 732
; || || || || || || || || || || || || || || || || || || || || ||
; DB 962 TTTCCTTGCCAGAAAAAATACCTTTTAACTTTATCTGTAATGTGTGCTTGAAAT 903
; || || || || || || || || || || || || || || || || || || || || ||
; QY 733 tctctctctcgtcagcttagcctgtgc 762
; || || || || || || || || || || || || || || || || || || || || ||
; DB 902 TTTCACCTGCTTCAGTCTTGCCCTGTTTC 873
; || || || || || || || || || || || || || || || || || || || || ||
;
; RESULT 15
; PCT-US95-13661-3/c
; Sequence 3, Application PC/TUS9513661
; GENERAL INFORMATION:
; APPLICANT: Beach, David H.
; APPLICANT: Galationov, Konstantin
; APPLICANT: Jessus, Catherine
; TITLE OF INVENTION: Interactions between Raf Proto-Oncogenes
; TITLE OF INVENTION: and CDC25 Phosphatases, and Uses Related Thereeto
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(Text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13661
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: CSI-002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2062 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 211..1631
; PCT-US95-13661-3
;
;
; Query Match 3.4%; Score 30.8; DB 5; Length 2062;
; Best Local Similarity 58.9%; Pred. No. 5.8;
; Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
;
; QY 673 ttccagtgctctcaaaagcatgttttcaagatcatttggttgctctctctagtg 732
; || || || || || || || || || || || || || || || || || || || || ||
; DB 962 TTTCCTTGCCAGAAAAAATACCTTTTAACTTTATCTGTAATGTGTGCTTGAAAT 903
; || || || || || || || || || || || || || || || || || || || || ||
; QY 733 tctctctctcgtcagcttagcctgtgc 762
; || || || || || || || || || || || || || || || || || || || || ||
; DB 902 TTTCACCTGCTTCAGTCTTGCCCTGTTTC 873
; || || || || || || || || || || || || || || || || || || || || ||

```

Search completed: September 16, 2002, 03:27:01
Job time: 9344 sec
